PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A61K 39/395, C12Q 1/68, G01N 33/53, C12N 15/00, 15/09, 15/11, 15/64

A1

(11) International Publication Number: WO 96/34627

(43) International Publication Date: 7 November 1996 (07.11.96)

(21) International Application Number: PCT/US96/06231

(22) International Filing Date: 2 May 1996 (02.05.96)

(30) Priority Data:
08/435,454 5 May 1995 (05.05.95) US

(71) Applicant: THE WISTAR INSTITUTE OF ANATOMY AND BIOLOGY [US/US]; 36th and Spruce Streets, Philadelphia, PA 19104-4268 (US).

(72) Inventors: PRENDERGAST, George, C.; 352 Birdsong Way, Doylestown, PA 18901 (US). SAKAMURO, Daitoku; International House #501, 3701 Chestnut Street, Philadelphia, PA 19104 (US).

(74) Agents: KODROFF, Cathy, A. et al.; Howson and Howson, Spring House Corporate Center, P.O. Box 457, Spring House, PA 19477 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

**Published**

*With international search report.*

*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: MURINE AND HUMAN BOX-DEPENDENT MYC-INTERACTING PROTEIN (BIN1) AND USES THEREFOR

(57) Abstract

A partial murine cDNA clone, a human cDNA clone, and a partial human genomic clone, each encoding a Box-dependent myc-interacting polypeptide termed Bin1, are provided. Also provided are methods of using the nucleic acid sequences, polypeptides, and antibodies directed against same in the diagnosis and treatment of cancers and hyperplastic disease states.

## *FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | |
|---|---|
| AM | Armenia |
| AT | Austria |
| AU | Australia |
| BB | Barbados |
| BE | Belgium |
| BF | Burkina Faso |
| BG | Bulgaria |
| BJ | Benin |
| BR | Brazil |
| BY | Belarus |
| CA | Canada |
| CF | Central African Republic |
| CG | Congo |
| CH | Switzerland |
| CI | Côte d'Ivoire |
| CM | Cameroon |
| CN | China |
| CS | Czechoslovakia |
| CZ | Czech Republic |
| DE | Germany |
| DK | Denmark |
| EE | Estonia |
| ES | Spain |
| FI | Finland |
| FR | France |
| GA | Gabon |
| GB | United Kingdom |
| GE | Georgia |
| GN | Guinea |
| GR | Greece |
| HU | Hungary |
| IE | Ireland |
| IT | Italy |
| JP | Japan |
| KE | Kenya |
| KG | Kyrgystan |
| KP | Democratic People's Republic of Korea |
| KR | Republic of Korea |
| KZ | Kazakhstan |
| LI | Liechtenstein |
| LK | Sri Lanka |
| LR | Liberia |
| LT | Lithuania |
| LU | Luxembourg |
| LV | Latvia |
| MC | Monaco |
| MD | Republic of Moldova |
| MG | Madagascar |
| ML | Mali |
| MN | Mongolia |
| MR | Mauritania |
| MW | Malawi |
| MX | Mexico |
| NE | Niger |
| NL | Netherlands |
| NO | Norway |
| NZ | New Zealand |
| PL | Poland |
| PT | Portugal |
| RO | Romania |
| RU | Russian Federation |
| SD | Sudan |
| SE | Sweden |
| SG | Singapore |
| SI | Slovenia |
| SK | Slovakia |
| SN | Senegal |
| SZ | Swaziland |
| TD | Chad |
| TG | Togo |
| TJ | Tajikistan |
| TT | Trinidad and Tobago |
| UA | Ukraine |
| UG | Uganda |
| US | United States of America |
| UZ | Uzbekistan |
| VN | Viet Nam |

# MURINE AND HUMAN BOX-DEPENDENT MYC-INTERACTING PROTEIN (BIN1) AND USES THEREFOR

This invention was made with financial assistance from the National Institutes of Health Grant No. 5-P30-CA-10815-28. The United States government has certain rights in this invention.

## Field of the Invention

This invention relates generally to cancer diagnosis and therapy, and more specifically, to cancers associated with the *Myc* oncoprotein.

## Background of the Invention

*Myc* is a transcription factor and key cell growth regulator that is frequently deregulated in human malignancy, notably Burkitt's and T cell lymphomas, where *myc* genes suffer chromosomal translocation. In colon and lung carcinomas, *myc* genes are amplified [M.D. Cole, *Ann. Rev. Genet.*, *20*:361-384 (1986)]. Paradoxically, under certain conditions *myc* can induce apoptosis, a regulated cell suicide process [D.S. Askew et al, *Oncogene*, *6*:1915-1922 (1991); G.I. Evan et al, *Cell*, *69*:119-128 (1992)]. However, loss or suppression of apoptosis is an important step in the malignant conversion of human tumors containing deregulated *myc* oncogenes, including, prominently, prostate carcinoma [T. G. Strohmeyer et al, *J. Urol.*, *151*:1479-1497 (1994)].

There remains a need in the art for compositions and methods of regulating a deregulated *Myc* protein and of exploiting and/or diagnosing its apoptotic potential.

## Brief Description of the Drawings

Fig. 1 is a partial murine cDNA sequence SEQ ID NO:1 and the murine Bin1 polypeptide encoded thereby SEQ ID NO:2.

Fig. 2A-2C is a human cDNA sequence SEQ ID NO:3 and the human Bin1 polypeptide encoded thereby SEQ ID NO:4.

Fig. 3A is a bar chart illustrating the selective requirement of the Myc-binding domain (MBD) for Myc inhibition, as described in Example 7. The data represent three to seven trials for each transfection. The data are depicted as the percent of foci induced by oncogenes and vector, as appropriate.

Fig. 3B is a bar chart illustrating the dominant inhibitory activity of MBD.

Fig. 4 is a bar chart illustrating that Bin1 vectors selectively inhibit colony formation in HepG2 cells lacking endogenous expression. The data are depicted as the percentage of colonies obtained with empty vector.

Summary of the Invention

In one aspect, the present invention provides a partial murine cDNA clone of a Box-dependent myc-interacting polypeptide 1 (Bin1), formerly referred to as c-Myc interacting peptide (MIP or MIP-99), SEQ ID NO:1, and the polypeptide encoded thereby, SEQ ID NO:2.

In another aspect, the present invention provides a human Bin1 cDNA clone, SEQ ID NO:3, and the human polypeptide encoded thereby, SEQ ID NO:4.

In yet another aspect, the present invention provides a vector comprising a mammalian nucleic acid sequence encoding a Bin1 protein and a host cell transformed by such a vector. Alternatively, this vector may be used in gene therapy applications.

In still another aspect, the invention provides an oligonucleotide probe comprising a nucleic acid sequence as defined herein. Also provided is an antibody raised against a Bin1 protein or peptide thereof.

In yet a further aspect, the present invention provides a diagnostic reagent for breast or liver cancer, or deficient Bin1 production, comprising an oligonucleotide probe or an antibody of the invention.

Further provided is a therapeutic reagent comprising a polypeptide, anti-idiotype antibody, or gene therapy vector of the invention.

Still another aspect of the invention provides a method of treating breast or liver cancer by administering a therapeutic reagent of the invention.

Other aspects and advantages of the present invention are described further in the following detailed description of the preferred embodiments thereof.

Detailed Description of the Invention

The present invention provides novel, isolated, nucleic acid sequences which encode novel proteins which interact with c-Myc and bind thereto, fragments of these sequences and antibodies developed thereto. The nucleic acid sequences, protein sequences and antibodies are useful in the detection, diagnosis and treatment of cancers or other disorders associated with deregulation, deficiency or amplification of the c-*myc* oncogenes. Further, when a Box-dependent myc-interacting polypeptide 1 (called Bin1) of this invention binds to c-Myc, the binding appears to regulate the c-Myc and result in tumor suppression, by inhibiting cell growth and/or facilitating apoptosis (programmed cell death). The Bin1 gene has several other features suggesting it is a tumor suppressor gene. First, Bin1 inhibits Myc-dependent malignant cell transformation. Second, Bin1 is structurally related to RVS167, a negative regulation of the cell division cycle in the yeast *Saccharomyces cerevisiae* [F. Bauer et al, Mol. Cell. Biol., 13:5070-5084 (1993)]. Third, Southern analysis of the Bin1 gene

reveals that it is mutated in a significant portion of human liver carcinoma cell lines. Fourth, Northern analysis indicates that expression of Bin1 RNA is lost in human liver and breast carcinoma cell lines. Fifth, chromosomal mapping has identified Bin1's location at 2q14, a frequent site of deletion in metastatic prostate cancers [W. Isaacs, Johns Hopkins Medical School, personal communication] and radiation-induced leukemias [I. Hayata et al, Cancer Res., 43:367-373 (1983)]. All of these features support the assignment of Bin1 as a tumor suppressor gene, similar to the breast cancer gene BRCA1, and the genes encoding p53 and the Rb retinoblastoma protein, which are negative regulators of cell growth that are observed to be mutated and/or unexpressed in human cancer cells. These aspects of the invention are discussed in more detail below.

I. Nucleic Acid Sequences

The present invention provides mammalian nucleic acid sequences encoding a Box-dependent myc-interacting polypeptide 1, termed herein Bin1. The nucleic acid sequences of this invention are isolated from cellular materials with which they are naturally associated. In one embodiment, a Bin1 nucleic acid sequence is selected from all or part of the partial murine cDNA clone, SEQ ID NO: 1. In another embodiment, a Bin1 nucleic acid sequence is selected from all or part of a human cDNA clone, SEQ ID NO: 3. In yet another embodiment, the present invention provides a partial Bin1 genomic sequence, SEQ ID NO: 6. However, the present invention is not limited to these nucleic acid sequences.

Given the sequences of SEQ ID NO: 1, SEQ ID NO: 3, and SEQ ID NO: 6, one of skill in the art can readily obtain the corresponding anti-sense strands of these cDNA and genomic sequences. Further, using known techniques,

one of skill in the art can readily obtain further genomic sequences corresponding to these cDNA sequences or the corresponding RNA sequences, as desired.

Similarly the availability of SEQ ID NOS: 1, 3 and 6 of this invention permits one of skill in the art to obtain other species Bin1 analogs, by use of the nucleic acid sequences of this invention as probes in a conventional technique, e.g., polymerase chain reaction. Allelic variants of these sequences within a species (i.e., nucleotide sequences containing some individual nucleotide differences from a more commonly occurring sequence within a species, but which nevertheless encode the same protein) such as other human variants of Bin1 SEQ ID NO: 3, may also be readily obtained given the knowledge of this sequence provided by this invention.

The present invention further encompasses nucleic acid sequences capable of hybridizing under stringent conditions [see, J. Sambrook et al, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory (1989)] to the sequences of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 6, their anti-sense strands, or biologically active fragments thereof. An example of a highly stringent hybridization condition is hybridization at 2XSSC at 65°C, followed by a washing in 0.1XSSC at 65°C for an hour. Alternatively, an exemplary highly stringent hybridization condition is in 50% formamide, 4XSSC at 42°C. Other, moderately high stringency conditions may also prove useful, e.g. hybridization in 4XSSC at 55°C, followed by washing in 0.1XSSC at 37°C for an hour. Alternatively, an exemplary moderately high stringency hybridization condition is in 50% formamide, 4XSSC at 30°C.

Also encompassed within this invention are fragments of the above-identified nucleic acid sequences. Preferably, such fragments are characterized by encoding

a biologically active portion of Bin1, e.g., an epitope. Generally, these oligonucleotide fragments are at least 15 nucleotides in length. However, oligonucleotide fragments of varying sizes may be selected as desired. Such fragments may be used for such purposes as performing the PCR, e.g., on a biopsied tissue sample. For example, one fragment which is anticipated to be particularly useful is the Src homology 3 (SH3) domain, which is located at about nucleotides 891-1412 of SEQ ID NO: 3 (which encode amino acid residues 278-451 of SEQ ID NO: 4). Preliminary data has indicated this domain may be useful in blocking apoptosis. Other useful fragments include about nucleotides 813-854 of SEQ ID NO: 3 (encoding a nuclear localization signal, amino acid residues about 252-265 of SEQ ID NO: 4), nucleotides about 867-908 (a Myc-binding domain or MBD amino acids 270-283). Other fragments and other uses of such fragments are discussed in more detail below.

The nucleotide sequences of the invention may be isolated by conventional uses of polymerase chain reaction or cloning techniques such as those described in obtaining the murine and human sequences, described below. Alternatively, these sequences may be constructed using conventional genetic engineering or chemical synthesis techniques.

According to the invention, the nucleic acid sequences [SEQ ID NO: 1, SEQ ID NO: 3, and SEQ ID NO: 6] may be modified. Utilizing the sequence data in these figures and in the sequence listing, it is within the skill of the art to obtain other polynucleotide sequences encoding the proteins of the invention. Such modifications at the nucleic acid level include, for example, modifications to the nucleotide sequences which are silent or which change the amino acids, e.g. to

improve expression or secretion. Also included are allelic variations, caused by the natural degeneracy of the genetic code.

Also encompassed by the present invention are mutants of the Bin1 gene provided herein. Such mutants include amino terminal, carboxy terminal or internal deletions which are useful as dominant inhibitor genes. Such a truncated, or deletion, mutant may be expressed for the purpose of inhibiting the activity of the full-length or wild-type gene. For example, it has been found that expression of the partial murine Bin1 provided herein [SEQ ID NO:2] acts in a dominant inhibitory manner to suppress normal Bin1 activity. Expression of this protein is described in Example 4 below. Another mutant encodes Bin1 deleted in the region encoding the MBD domain (amino acid residues 270-383 of SEQ ID NO: 4).

The invention further provides the complete human Bin1 gene, which has been cloned as a 35-45 kb contiguous sequence from a lambda phage genomic library. The DNA sequence of approximately 19 kb (about the 3' half) of the approximately 40 kb Bin1 gene has been determined [SEQ ID NO: 6]. More detailed discussion of the Bin1 genomic sequence is provided in Example 3. The exon-intron junction sequences derived are desirable for applying PCR technology to identify mutations in DNA derived from tumor biopsies, using techniques similar to those applied to sequences derived from other tumor suppressor genes (e.g., p53 and BRCA1). The sequenced region of the Bin1 gene spans regions previously found to be rearranged in liver and cervix carcinoma cell lines, making it possible to map deletions and possible mutations in primary human tumor DNA by PCR technology. Using the genomic clones, the human Bin1 gene has been mapped to chromosome 2q14, a region frequently deleted in prostate carcinoma and in radiation-induced malignancies.

These nucleic acid sequences are useful for a variety of diagnostic and therapeutic uses. Advantageously, the nucleic acid sequences are useful in the development of diagnostic probes and antisense probes for use in the detection and diagnosis of conditions characterized by deregulation or amplification of c-myc. The nucleic acid sequences of this invention are also useful in the production of mammalian, and particularly, murine and human Bin1 proteins.

## II. Protein Sequences

The present invention also provides mammalian Bin1 polypeptides or proteins. These proteins are free from association with other contaminating proteins or materials with which they are found in nature. In one embodiment, the invention provides a partial murine Bin1 [SEQ ID NO:2] polypeptide of 135 amino acids having a predicted molecular weight (MW) of 13,688. In another embodiment, the invention provides a full-length human Bin1 [SEQ ID NO:4] of 451 amino acids with an estimated MW of 50,048. The apparent MW of human Bin1 on sodium dodecyl sulfate polyacrylamide (SDS-PA) gels is approximately 67 kD.

Comparisons of the Bin1 amino acid sequence to the DNA database were performed using the search algorithm BLAST [S.F. Altschul et al, J. Mol. Biol., 215:403-410 (1990)]. Using the complete sequence to search the database, two known genes were identified which had highly significant similarity to the terminal regions of Bin1 ($p < 10^8$). The first gene was amphiphysin, a neuronal protein of unknown function which is the putative autoimmune antigen in breast cancer-associated Stiff-Man syndrome [F. Folli et al, N. Eng. J. Med., 328:546-551 (1993)], a paraneoplastic disorder that clinically presents in a fraction of breast cancer

patients. The second gene was RVS167, a negative regulator of the cell division cycle in S. cerevisiae. The region of the most extensive similarity between amphiphysin and RVS167, approximately 50% and 25%, respectively, lies within residues 1-222 of Bin1 [SEQ ID NO:4]. Therefore, this N-terminal region of Bin1 has been termed herein the BAR domain (for Bin1/amphiphysin/RVS167-related domain). The extensive similarity of the BAR domains in these proteins suggest a common molecular function. Moreover, the relationship suggests roles for Bin1 in breast malignancy, where Myc is frequently involved, and in cell cycle regulation. Finally, since RVS167 is a negative regulator which is dispensible for cell growth but required for cell cycle exit, the similarity to RVS167 would be consistent with the likelihood that Bin1 is a tumor suppressor.

To gain additional insights into the molecular functions of Bin1, additional BLAST searches were performed with subsections of the Bin1 sequence. These searches identified several gene products which all function in regulation of cell cycle transit and/or chromosomal structure. Several additional relationships were revealed within the Bin1 BAR domain. These included pericentrin (30% identical; 46% similar; P<0.01), a centromere-binding protein required for proper chromosome organization during the cell cycle M phase; mitosin (24% identical; 48% similar; P=0.02), a protein implicated in transit through M phase; and SMC1 (21% identical; 43% similar; P=0.05), a yeast regulator of M phase chromosome segregation. In the scoring range where these similarities were observed, highly alpha helical regions of non-muscle myosin, tropomyosin, and the trp gene product were also found, suggesting that the BAR domain shares their highly helical structure. Between the C-terminal end of the BAR region and the nuclear

localization signal (NLS; amino acids 252-265, SEQ ID NO: 4) lies an additional Bin1 domain (amino acids 224-251, SEQ ID NO: 4), encoded by a single exon, which is not found in amphiphysin and RVS167 but which also contains motifs seen in proteins controlling cell cycle and chromosome structure. One ~10 amino acid motif is found in a functionally important region of the SV40 T antigen oncoprotein, while a second motif is seen in RED1, a yeast protein implicated in chromosome segregation. Proximal to these motifs is an additional motif which is similar to p93dis1, another yeast protein implicated in chromosome segregation. Taken together, these observations strengthen the likelihood that Bin1 participates in some aspect of cell cycle regulation and further suggests a role in chromosome structure control.

Further encompassed by this invention are fragments of the Bin1 polypeptides. Such fragments are desirably characterized by having Bin1 biological activity, including, e.g., the ability to interact with c-Myc. These fragments may be designed or obtained in any desired length, including as small as about 8 amino acids in length. Such a fragment may represent an epitope of the protein. One particularly desirable fragment is located at amino acid residues 270-383 of SEQ ID NO: 4, which is the c-Myc binding domain (MBD). Another desirable fragment is located at residues 278-451 of SEQ ID NO: 4 and is a Src homology 3 (SH3) domain. A third fragment is located at residues 223-251 of SEQ ID NO:4 and includes the T antigen/RED1/p93dis1 motifs discussed above. Yet another desirable fragment includes the BAR domain, located at amino acid residues 1-222 of SEQ ID NO:4. Finally, a fragment containing the nuclear localization domain located at amino acid residues 252 to about 265 of SEQ ID NO: 4, may also be desirable.

Also included in the invention are analogs, or modified versions, of the proteins provided herein. Typically, such analogs differ by only one to four codon changes. Examples include polypeptides with minor amino acid variations from the illustrated amino acid sequences of Bin1 (Figs. 1 and 2; SEQ ID NO:2 and 4); in particular, conservative amino acid replacements. Conservative replacements are those that take place within a family of amino acids that are related in their side chains and chemical properties. Also provided are homologs of the proteins of the invention which are characterized by having at least 85% homology with SEQ ID NO:2 or SEQ ID NO:4. It has previously determined that the murine and human Bin1 (in partial) are about 88.5% identical.

Additionally, the Bin1 proteins [SEQ ID NO:2 and 4] of the invention may be modified, for example, by truncation at the amino or carboxy termini, by elimination or substitution of one or more amino acids, or by any number of now conventional techniques to improve production thereof, to enhance protein stability or other characteristics, e.g. binding activity or bioavailability, or to confer some other desired property upon the protein.

## III. <u>Expression</u>

### A. <u>In Vitro</u>

To produce recombinant Bin1 proteins of this invention, the DNA sequences of the invention are inserted into a suitable expression system. Desirably, a recombinant molecule or vector is constructed in which the polynucleotide sequence encoding Bin1 is operably linked to a heterologous expression control sequence permitting expression of the murine or human Bin1 protein. Numerous types of appropriate expression

vectors are known in the art for mammalian (including human) protein expression, by standard molecular biology techniques. Such vectors may be selected from among conventional vector types including insects, e.g., baculovirus expression, or yeast, fungal, bacterial or viral expression systems. Other appropriate expression vectors, of which numerous types are known in the art, can also be used for this purpose.

Methods for obtaining such expression vectors are well-known. See, Sambrook et al, Molecular Cloning. A Laboratory Manual, 2d edition, Cold Spring Harbor Laboratory, New York (1989); Miller et al, Genetic Engineering, 8:277-298 (Plenum Press 1986) and references cited therein.

Suitable host cells or cell lines for transfection by this method include mammalian cells, such as Human 293 cells, Chinese hamster ovary cells (CHO), the monkey COS-1 cell line or murine 3T3 cells derived from Swiss, Balb-c or NIH mice may be used. Another suitable mammalian cell line is the CV-1 cell line. Still other suitable mammalian host cells, as well as methods for transfection, culture, amplification, screening, and product production and purification are known in the art. [See, e.g., Gething and Sambrook, Nature, 293:620-625 (1981), or alternatively, Kaufman et al, Mol. Cell. Biol., 5(7):1750-1759 (1985) or Howley et al, U. S. Patent 4,419,446].

Similarly bacterial cells are useful as host cells for the present invention. For example, the various strains of E. coli (e.g., HB101, MC1061, and strains used in the following examples) are well-known as host cells in the field of biotechnology. Various strains of B. subtilis, Pseudomonas, other bacilli and the like may also be employed in this method.

Many strains of yeast cells known to those skilled in the art are also available as host cells for expression of the polypeptides of the present invention. Other fungal cells may also be employed as expression systems.

Alternatively, insect cells such as Spodoptera frugipedera (Sf9) cells may be used.

Thus, the present invention provides a method for producing a recombinant Bin1 protein which involves transfecting a host cell with at least one expression vector containing a recombinant polynucleotide encoding a Bin1 protein under the control of a transcriptional regulatory sequence, e.g. by conventional means such as electroporation. The transfected host cell is then cultured under suitable conditions that allow expression of the Bin1 protein. The expressed protein is then recovered, isolated, and optionally purified from the culture medium (or from the cell, if expressed intracellularly) by appropriate means known to one of skill in the art.

For example, the proteins may be isolated in soluble form following cell lysis, or may be extracted using known techniques, e.g., in guanidine chloride. If desired, the Bin1 proteins of the invention may be produced as a fusion protein. For example, it may be desirable to produce Bin1 fusion proteins, to enhance expression of the protein in a selected host cell, to improve purification, or for use in monitoring the presence of Bin1 in tissues, cells or cell extracts. Suitable fusion partners for the Bin1 proteins of the invention are well known to those of skill in the art and include, among others, $\beta$-galactosidase, glutathione-S-transferase, and poly-histidine.

B. In Vivo

Alternatively, where it is desired that the Bin1 protein be expressed in vivo, e.g., for gene therapy purposes, an appropriate vector for delivery of Bin1, or fragment thereof (such as the SH3 domain), may be readily selected by one of skill in the art. Exemplary gene therapy vectors are readily available from a variety of academic and commercial sources, and include, e.g., adeno-associated virus [International patent application No. PCT/US91/03440], adenovirus vectors [M. Kay et al, Proc. Natl. Acad. Sci. USA, 91:2353 (1994); S. Ishibashi et al, J. Clin. Invest., 92:883 (1993)], or other viral vectors, e.g., various poxviruses, vaccinia, etc. Methods for insertion of a desired gene, e.g. Bin1, and obtaining in vivo expression of the encoded protein, are well known to those of skill in the art.

IV. Antisera and Antibodies

The Bin1 proteins of this invention are also useful as antigens for the development of anti-Bin1 antisera and antibodies to Bin1 or to a desired fragment of a Bin1 protein. Specific antisera may be generated using known techniques. See, Sambrook, cited above, Chapter 18, generally, incorporated by reference. Similarly, antibodies of the invention, both polyclonal and monoclonal, may be produced by conventional methods, including the Kohler and Milstein hybridoma technique, recombinant techniques, such as described by Huse et al, Science, 246:1275-1281 (1988), or any other techniques known to the art. For example, rabbit polyclonal antisera was developed and recognizes an epitope(s) between amino acid residues 190-250 of SEQ ID NO: 4. This antisera has been found to be human-specific. Since amino acids 190-250 are outside the MBD and SH3 domain, for experimental needs two additional antisera have been

raised to these regions. The immunogens included human Bin1 amino acids 270-383 (MBD) [SEQ ID NO: 4] or amino acids 278-451 (SH3) [SEQ ID NO: 4]. Each antisera has been shown to recognize the appropriate domain by immunoprecipitation.

Additionally, six (6) Bin1-specific monoclonal antibodies have been characterized, termed 99-D through 99-I. The approximate location of the epitopes within Bin1 for each antibody has been mapped. MAb 99D recognizes an epitope within amino acids 190-250 [SEQ ID NO: 4]; MAbs 99F-99I recognize epitopes within the NLS (amino acids 252-261 [SEQ ID NO: 4]); MAb 99E recognizes a complex epitope requiring amino acids 190-250 and amino acids 263-397 [SEQ ID NO: 4]. Each antibody has been isotyped and demonstrated to work in immunoprecipitation, Western blotting, and immunohistochemistry methodology. Particularly, MAb 99D and MAb 99F are IgG2b isotypes; MAbs 9E, 99G and 99H are IgG1 isotypes. Further, MAbs 99D and 99F have been determined to be useful for immunohistochemistry with sectioned biopsy tissue and tissue culture cells, and are therefor likely to be useful for clinical applications to analyze tumor biopsies. 99D recognizes a nuclear protein present in all normal cells examined so far but missing in carcinoma cells previously demonstrated to lack Bin1 RNA. 99F has been determined to specifically recognize a cytoplasmic form of Bin1 which is induced following muscle differentiation in an <u>in vitro</u> model system which is described below. 99D recognizes both the cytoplasmic as well as the nuclear forms of Bin1. 99D has been determined to be effective for detecting Bin1 protein by standard Western methodology in nonionic detergent lysates of a wide variety of tissues and tissue culture cells. 99D and 99F also have been shown to recognize both murine and human Bin1 polypeptides.

Also encompassed within this invention are humanized and chimeric antibodies. As used herein, a humanized antibody is defined as an antibody containing murine complementary determining regions (CDRs) capable of binding to Bin1 or a fragment thereof, and human framework regions. These CDRs are preferably derived from a murine monoclonal antibody (MAb) of the invention. As defined herein, a chimeric antibody is defined as an antibody containing the variable region light and heavy chains, including both CDR and framework regions, from a Bin1 MAb of the invention and the constant region light and heavy chains from a human antibody. Methods of identifying suitable human framework regions and modifying a MAb of the invention to contain same to produce a humanized or chimeric antibody of the invention, are well known to those of skill in the art. See, e.g., E. Mark and Padlin, "Humanization of Monoclonal Antibodies", Chapter 4, <u>The Handbook of Experimental Pharmacology</u>, Vol. 113, The Pharmacology of Monoclonal Antibodies, Springer-Verlag (June, 1994). Other types of recombinantly-designed antibodies are also encompassed by this invention.

Further provided by the present invention are anti-idiotype antibodies (Ab2) and anti-anti-idiotype antibodies (Ab3). Ab2 are specific for the target to which anti-Bin1 antibodies of the invention bind and Ab3 are similar to Bin1 antibodies (Ab1) in their binding specificities and biological activities [see, e.g., M. Wettendorff et al., "Modulation of anti-tumor immunity by anti-idiotypic antibodies." In <u>Idiotypic Network and Diseases</u>, ed. by J. Cerny and J. Hiernaux J, Am. Soc. Microbiol., Washington DC: pp. 203-229, (1990)]. These anti-idiotype and anti-anti idiotype antibodies may be produced using techniques well known to those of skill in the art. Such anti-idiotype antibodies (Ab2) can bear

the internal image of the c-Myc and bind to it in much the same manner as Bin1 and are thus useful for the same purposes as Bin1.

In general, polyclonal antisera, monoclonal antibodies and other antibodies which bind to Bin1 as the antigen (Ab1) are useful to identify epitopes of Bin1, to separate Bin1 from contaminants in living tissue (e.g., in chromatographic columns and the like), and in general as research tools and as starting material essential for the development of other types of antibodies described above. Anti-idiotype antibodies (Ab2) are useful for binding c-Myc and thus may be used in the treatment of cancers in which c-Myc is part of a biochemical cascade of events leading to tumor formation. The Ab3 antibodies may be useful for the same reason the Ab1 are useful. Other uses as research tools and as components for separation of c-Myc from other contaminant of living tissue, for example, are also contemplated for these antibodies.

V. Diagnostic Reagents and Methods

Advantageously, the present invention provides reagents and methods useful in detecting and diagnosing abnormal levels of Bin1, and particularly deficiencies or excess production thereof, in a patient. As defined herein, a deficiency of Bin1 is defined as an inadequate amount of Bin1 to compensate for the levels of c-Myc in a patient. Conditions associated with deficiencies of Bin1 include a variety of cancers, e.g., breast cancer, liver cancer and colon cancer, and hyperplastic disease states, e.g., benign prostate hyperplasia.

Thus, the proteins, protein fragments, antibodies, and polynucleotide sequences (including anti-sense polynucleotide sequences and oligonucleotide fragments), and Bin1 antisera and antibodies of this invention may be

useful as diagnostic reagents. These reagents may optionally be labelled using diagnostic labels, such as radioactive labels, colorimetric enzyme label systems and the like conventionally used in diagnostic or therapeutic methods. Alternatively, the N- or C-terminus of Bin1 or a fragment thereof may be tagged with a viral epitope which can be recognized by a specific antisera. The reagents may be used to measure abnormal Bin1 levels in selected mammalian tissue in conventional diagnostic assays, e.g., Southern blotting, Northern and Western blotting, polymerase chain reaction (PCR), reverse transcriptase (RT) PCR, immunostaining, and the like. For example, in biopsies of tumor tissue, loss of Bin1 expression in tumor tissue could be directly verified by RT-PCR or immunostaining. Alternatively, a Southern analysis, genomic PCR, or fluorescence in situ hybridization (FISH) may be performed to confirm Bin1 gene rearrangement.

In one example, as diagnostic agents the polynucleotide sequences may be employed to detect or quantitate normal Bin1. The selection of the appropriate assay format and label system is within the skill of the art and may readily be chosen without requiring additional explanation by resort to the wealth of art in the diagnostic area.

Thus the present invention provides methods for the detection of disorders characterized by insufficient Bin1 levels. Currently, it is anticipated that antibodies of the invention, such as 99D and 99F, which have been found to be able to withstand the conditions necessary for tissue fixation, will be particularly useful for biopsies. However, the protein, antibody, antisera or polynucleotide reagents of the invention are expected to be similarly useful in the following methods. The methods involve contacting a selected mammalian tissue,

e.g., a biopsy sample or other cells, with the selected reagent, protein, antisera antibody or DNA sequence, and measuring or detecting the amount of Bin1 present in the tissue in a selected assay format based on the binding or hybridization of the reagent to the tissue.

VI. Therapeutic Compositions and Methods

Compositions and methods useful for the treatment of conditions associated with inadequate Bin1 levels are provided. As stated above, included among such conditions are liver, colon and breast cancers and hyperplastic disease states. Also provided are compositions and methods for inhibiting Bin1 activity in order to ameliorate a condition in which apoptosis is activated and Bin1 plays a role. Such conditions may include degenerative conditions, e.g., neurodegenerative diseases.

The therapeutic compositions of the invention may be formulated to contain an anti-idiotype antibody of the invention, or the Bin1 protein itself or a fragment thereof may be administered to mimic the effect of normal Bin1 and bind c-Myc, thereby preventing its cancer causing function. For example, one particularly useful protein may be the Bin1 SH3 domain (amino acids 378-451 of SEQ ID NO: 4). These compositions may contain a pharmaceutically acceptable carrier. Suitable carriers are well known to those of skill in the art and include, for example, saline. Alternatively, such compositions may include conventional delivery systems into which protein of the invention is incorporated. Optionally, these compositions may contain other active ingredients, e.g., chemotherapeutics.

Still another method involves the use of the Bin1 polynucleotide sequences for gene therapy. In the method, the Bin1 sequences are introduced into a suitable

vector for delivery to a cell containing a deficiency of Bin1 and/or to block tumor growth. By conventional genetic engineering techniques, the Bin1 gene sequence may be introduced to mutate the existing gene by recombination or to replace an inactive or missing gene.

The dose, timing and mode of administration of these compositions may be determined by one of skill in the art. Such factors as the age, condition, and the level of the Bin1 deficiency detected by the diagnostic methods described above, may be taken into account in determining the dose, timing and mode of administration of the therapeutic compositions of the invention. Generally, where treatment of an existing cancer or hyperplastic state is indicated, a therapeutic composition of the invention is preferably administered in a site-directed manner and is repeated as needed. Such therapy may be administered in conjunction with conventional therapies, including radiation and/or chemotherapeutic treatments.

The following examples illustrate the isolation and use of the Bin1 sequences of the invention. These examples are illustrative only and do not limit the scope of the invention.

Example 1 - Identification and Characterization of Bin1

A. Murine Bin1 cDNA

A yeast two hybrid approach [Fields, S. and O. Song., Nature, 340:245-6 (1989)] was used to screen for Myc-interacting proteins (Bin1) in a murine embryonic cDNA library. The cDNA library was derived from day 10.5 mouse embryonic RNA [A. B. Vojtek et al, Cell, 74: 205-214 (1993)]. This system takes advantage of the modular nature of transcription factors, whose DNA-binding and transcriptional activating components can be assembled in trans by interacting protein (IP) domains derived from other polypeptides. A previously described

two hybrid system [Vojtek et al, cited above] and a 16 amino acid nontransactivating polypeptide derived from the human c-Myc "Myc box 1" (MB1) region [Prendergast, G.C. and E.B. Ziff, Trends in Genet., 8: 91-96.3 (1992)] EDIWKKFELLPTPPLS (human c-Myc amino acids 47-62) [SEQ ID NO:5], were used as "bait" in the screen.

Briefly, the "bait" plasmid contained a TRP1 marker and a LexA-MB1 fusion protein as the DNA binding component, and the cDNA library vector, pVP16, contained a LEU2 marker and the herpes simplex virus VP16 protein as the transcriptional transactivator fused to the cDNA library inserts. cDNA synthesized from the 10.5 day murine embryo RNA was size-selected by random DNaseI treatment to ~0.05 kb, treated with Klenow enzyme, NotI linked, and subcloned into pVP16. This cDNA library was designed to express protein modules whose interactions might be occluded in full-length polypeptides. The yeast strain L40 (MATa trp1-901 leu2-3,112 LYS2::(lexAop)4-HIS3 URA3::(lexAop)8-lacZ) served as the host for the two hybrid screen [see, Vojtek et al, cited above].

An L40 derivative expressing the MB1 "bait" was transfected with the cDNA library and approximately 3 x $10^7$ TRP+LEU+ transformants were examined in the primary screen, 300-400 of which were also the HIS+LacZ+ phenotype, which is diagnostic for interaction between the "bait" and library components [Vojtek et al, cited above]. The clones were cured of the original "bait" plasmid by standard methods [Guthrie, C. and G.R. Fink, eds., Guide to Yeast Genetics and Molecular Biology, Meth. Enzymol., 194, Academic Press: New York (1991)]. One hundred clones cured of the bait plasmid were tested for interaction by a mating strategy with a set of test baits.

The test "baits" included the original lexA-MB1 peptide construct, a set of negative controls that

included no insert, lamin [A. B. Vojtek et al, Cell, 74:205-214 (1993)], the small GTP-binding protein RhoB [D. Jahner, Mol. Cell. Biol., 11:3682-3690 (1991)], the peptide FTRHPPVLTPPDQEVI [SEQ ID NO: 7] derived from rat protein kinase Cβ2, a mutant MB1 peptide, a similarly sized but nonspecific peptide derived from protein kinase C, or lamin. The protein kinase C (PKC) peptide contained a phosphorylation site structurally analogous to the MB1 T58 phosphorylation site, which is recognized by glycogen synthase kinase-3 (GSK-3), a kinase present in yeast. The PKC peptide was designed to control for binding proteins that might non-specifically interact with phosphooligopeptides (e.g., peptidases, kinases, phosphatases). MB1 specificity was reproducibly exhibited by 14/99 of the original yeast clones.

cDNA library plasmids were shuttled from the desired clones to E. coli [Guthrie et al, cited above] and the DNA sequence of the inserts was determined. All clones contained related or identical sequences of approximately 0.4 kb containing an open reading frame (ORF) of 135 amino acids encoding a Myc-interacting polypeptide, termed Bin1 [SEQ ID NO:2], which exhibited specificity for Myc.

B. Bacterial Expression of murine Bin1 polypeptide [SEQ ID NO:2] as a soluble GST fusion protein

To study the association of the 135 aa murine Bin1 polypeptide [SEQ ID NO:2] with Myc in vitro, the ~0.4 kb cDNA [SEQ ID NO:1] was expressed as a glutathione-S-transferase (GST) fusion protein and used in binding assays with $^{35}$S-methionine-labeled in vitro translated (IVT) proteins. The binding experiments were configured essentially as described in A. K. Rustgi et al, Nature, 352:541-544 (1991).

To construct the GST fusion protein, the murine cDNA insert on a ClaI-EcoRI fragment was substituted for a similar fragment in pE47 [C. Murre et al, Cell, 56:777-783 (1989)], making pATG-99. The pATG-99 ORF included an initiator methionine, added a 15 amino acid N-terminal extension (3 amino acids from E47 and 12 amino acids from VP16) to the 135 residue clone #99 ORF, and retained the translational termination site derived from the two hybrid vector. Expression of the ATG99 polypeptide was confirmed by in vitro translation from pATG-99. The pATG-99 insert was then subcloned into pGEX-2T (Pharmacia) and the recombinant plasmid introduced into E. coli. GST-99 polypeptide was expressed and purified from E. coli cell extracts on glutathione-Sepharose (Pharmacia), using protocols supplied by the vendor.

Twenty (20) $\mu$l (~0.5 $\mu$g) of purified GST-99 protein was analyzed on an SDS-PA gel fixed and stained with Coomassie Blue. The apparent molecular weight (MW) of the Bin1 component of the fusion (22 kD) is larger than the predicted MW (14 kD) but is consistent with the apparent MW of in vitro translated murine Bin1 [SEQ ID NO: 2].

C. In Vitro Association of Myc and Bin1 [SEQ ID NO: 2]

[$^{35}$S]-methionine labeled c-Myc polypeptides were generated by IVT using TNT rabbit reticulocyte lysates (Promega). Expression plasmids included CMV Hm [G. C. Prendergast et al, Cell, 65:395-407 (1991)]; CMV Hm subclones containing MB1 deletion amino acids 49-101 [J. Stone et al, Mol. Cell. Biol., 7:1697-1709 (1987)]; MB2 deletion amino acids 120-140 [L. Li et al, EMBO J., 13:4070-4079 (1994)], or both deletions; the adenovirus E1A vectors p12S, p13S; and the SV40 large T antigen vector pTag [unpublished data]; and CMV-USF [L. Li, cited above].

Approximately 2.5 μg of GST or GST-99 and 10 μl of an IVT reaction were added to 0.5 ml binding buffer (10 mM TrisCl pH 7.5, 5 mM EDTA, 500 mM NaCl, 0.25% NP40) incubated 1 hr at 4°C on a nutator shaker, washed four times with binding buffer, and analyzed by SDS-PAGE and fluorography. c-Myc (but none of the other polypeptides produced by IVT) exhibited association with GST-99.

D. Association of Bin1 [SEQ ID NO:2] with TBP but not USF

[$^{35}$S]-labeled TBP and USF were generated by IVT and tested for GST-99 binding as in C. above. Reinforcing the notion that it might be involved in MB1 function in transcriptional regulation by Myc, Bin1 bound to TATA-binding protein [TBP, a critical component of the basal transcription apparatus]. Other polypeptides that were tested for GST-99 interaction and found to be negative included Max, cell cycle protein p107, transcription factor YY1, extracellular protein PAI-1, small GTP-binding protein RhoB, and empty-vector-associated products. Taken together, these findings argued that the association between GST-99 and Myc was both specific and physiologically relevant, since it depended upon the presence of the Myc boxes.

Example 2 - Isolation of Human Bin1 cDNA

BLAST searches of the complete DNA sequence database [GenBank] with the murine Bin1 sequence showed no strong similarities to known genes, but revealed an approximately 89% identity to an 289 bp uncharacterized human "expressed sequence tag". This finding suggested that Bin1 represented a novel gene conserved and expressed in humans.

Northern analysis of RNA from several human tissues using a murine Bin1 cDNA [SEQ ID NO: 1] as probe revealed a single RNA species of ~2.2 kb that was abundant in

skeletal tissue. A 1.95 kilobase human Bin1 cDNA was obtained from a human skeletal muscle λZAPII cDNA library (Stratagene, La Jolla, CA) by standard methods [Sambrook et al, cited above], using the murine Bin1 probe, i.e., by hybridization with [$^{32}$P]-labeled clone #99 insert and washing under low stringency conditions (2 x SSC 42°C). The complete sequence of this ~2.0 kb full-length cDNA, p99f, was determined [SEQ ID NO: 3] using the dideoxy method with Sequenase (US Biochemicals) and assembled and analyzed with MacVector software (IBI/Kodak). DNA database comparisons were performed using BLAST software. The subcloned cDNA contained a 451 amino acid ORF with approximately 88% identity to a C-terminal region of murine Bin1. The human ORF was therefore designated human Bin1 [SEQ ID NO:4].

Example 3 - Human Bin1 Gene Isolation, Structure and Regulation

A. DNA sequencing of the human Bin1 gene

Genomic clones of human Bin1 have been obtained. A 40 Kb contiguous sequence composed of five lambda phage genomic inserts has been assembled which contains the entire Bin1 gene. Approximately 15kb of the gene sequence is provided in SEQ ID NO: 6. Eight contiguous exons from the C-terminal BAR region to the SH3 domain have been identified. Three additional N-terminal BAR exons have also been identified. Five other exons identified by DNA sequence analysis algorithms appear in alternatively spliced RNAs found to be expressed exclusively in brain. With reference to the features information provided with respect to SEQ ID NO: 6, the nine exon sequences correspond to the following Bin1 cDNA sequences [SEQ ID NO: 3]: 623-655 (partial sequence of BAR region exon); 656-731 (3' BAR region exon); 732-814 (U1 region exon); 815-859 (NLS); 860-1004

(U2 region exon); 1005-1094 (5' MBD region exon); 1095-1205 (3' MBD region exon); 1206-1307 (5' SH3 domain region exon); 1308-1925 (3' SH3 domain/3' untranslated region [UTR] exon).

Using the genomic clones, the human Bin1 gene has been mapped to chromosome 2q14. This region is within a mid-2q locus that has been reported to be deleted in approximately 50% of metastatic prostate carcinomas. The region of the murine genome syntenic to human 2q14 has also been reported to be deleted in >90% of radiation induced leukemias and lymphomas. These data strengthen the previous assertion that Bin1 may be encoded by a novel tumor suppressor gene.

B. Increase in Bin1 Levels During Muscle and Neuronal Differentiation

Bin1 RNA has been found to be present in brain and muscle cells at 10- to 100-fold higher levels than other tissues, a feature shared with cell cycle kinase inhibitors (CKIs). Since these cells are postmitotic and Bin1 had been shown to block Myc's ability to induce cell cycle progression, it is possible that upregulation of Bin1 has a role in cell cycle exit associated with cell differentiation. To begin to assess this possibility, Bin1 expression was examined using in vitro model systems for differentiation of muscle cells (murine C2C12 premyoblast cells) [L. Silberstein et al, Cell, 46:1075-1081 (1986)] and neurons (rat PC12 pheochromocytoma cells) [L. A. Greene and A.S. Tischler, Proc. Natl. Acad. Sci. USA, 73:2424-2428 (1976)].

This analysis revealed that both Bin1 RNA and protein are regulated during cell differentiation. Bin1 RNA levels were increased following induction of cell differentiation in C2C12 or PC12 cells, by serum deprival or nerve growth factor (NGF) addition, respectively. In untreated PC12 cells, three transcripts of ~1.3, ~2.4,

and ~2.9 kb were noted. Within 5 days of NGF treatment the level of the ~2.9 kb RNA was increased several-fold, concomitant with neurite extension, while the level of the other two RNAs decreased to undetectable levels. The nature of the ~1.2 kb transcript, which was most abundant in untreated cells, was unclear but its unusually small size suggested the possibility that it was truncated due to mutation (PC12 was derived from a rat adrenal gland tumor). In C2C12 cells, a single ~2.4 kb transcript noted increased ~20-fold within 5 days of serum deprival, concomitant with myotube formation. These observations suggest that Bin1 may be involved in cell cycle regulation during neuronal and muscle cell differentiation.

C. <u>Identification of a larger Bin1 Polypeptide in Differentiated Muscle Cells</u>

Western analysis with the 99D antibody confirmed an increase in Bin1 expression and revealed the presence of a slightly larger Bin1 polypeptide generated 3 days post-induction. Levels of the smaller Bin1 polypeptide detected in undifferentiated cells was found to remain constant while the larger species increased dramatically. Indirect immunofluorescence using 99D antibody was used to examine the cell localization of Bin1 during C2C12 differentiation. Bin1 staining was found to change from a strictly nuclear pattern to whole cell pattern including the cytoplasm. The 99F antibody was found to detect only the larger polypeptide and stain only the cytoplasm (did not stain the nuclear protein). Thus, the larger Bin1 polypeptide induced during differentiation is completely confined to the cytoplasm. A dominant negative genetic approach is being taken to determine whether induction of the large Bin1 species is necessary for cell cycle exit during differentiation.

Example 4 - Construction of Mammalian Expression Vectors and Immunoprecipitation Techniques

Viral vectors for delivering Bin1 into insect, rodent and human cells have been developed for various purposes, including therapeutic purposes and to permit high-level Bin1 protein production and efficient gene transfer.

A. Baculoviral Vector

~1.6 kb EcoRI fragment containing the complete Bin1 coding region was inserted into the baculovirus recombination vector pVL1393 (Invitrogen, Inc., San Diego, CA), generating pBacBin. Sf9 insect cells were cotransfected with pBacBin and a plasmid encoding a defective baculovirus which cannot propagate. Rare recombination between these two plasmids in vivo leads to generation of a lytic recombinant baculovirus which can be propagated. Virus produced in cultures of cotransfected cells was propagated in mass Sf9 culture. Bin1 production was verified by Western analysis of NP40 lysates prepared 24 and 48 hr after infection of Sf9 cells infected with the BacBin virus, using 99D monoclonal antibody.

B. Adenoviral vector

The strategy and plasmid vector systems to produce recombinant adenovirus has been described [K. Kozarsky et al, Curr. Opin. Genet. Dev., 3:499-503 (1993)]. Similar to the approach taken to make baculoviral vectors, two plasmids are used which contain complementary regions which can homologously recombine in vivo. Recombinant virus is produced only in transfected cells where recombination has taken place. The plasmid pAdCMVpAT153 is used to introduce the gene of interest. pAdCMVpAT153 contains the left 6% of the adenovirus serotype 5 genome, modified such that the E1 region is replaced with a cytomegalovirus (CMV) early region

enhancer/promoter, multiple cloning site, and a G418 resistance gene cassette. Included in the cell transfection with this vector is a ~34 kb ClaI-digested fragment of adenovirus type 5 DNA that includes the remainder of the adenoviral genome. This fragment contains a mutation in the E3 region which ablates the immune response in adenovirus-infected animals [T. Ranheim et al, J. Virol., 67:2159-2167 (1993)]. This feature was incorporated into the recombinant virus to increase the persistence and therefore the potential efficacy of Bin1-based gene therapy approaches. The cell host for transfection is human 293 cells, an epithelial line which expresses the E1 region gene products required for propagation of recombinant adenoviruses.

The plasmid pAdenoBin was generated by inserting a ~1.6 kb EcoRI fragment containing the complete Bin1 coding region into the multiple cloning site of pAdCMVpAT153. 293 cells cotransfected with pAdenoBin and the ClaI-digested adenoviral DNA fragment were subjected to G418 selection and screening and purification by plaque assay (recombinant viruses are lytic in 293 cells). DNA isolated from a Bin1 virus identified in this manner will be validated by Southern analysis to confirm that the Bin1 cDNA is intact.

These vectors are particularly well suited for use in human therapies.

C. Moloney retroviral vector

A recombinant Bin1 retrovirus was generated using methods that have been described [N. Landau et al, J. Virol., 66:5110-5113 (1992)]. The Bin1 plasmid vector pSRαMSV-Bin1 was generated by inserting the ~1.6 kb EcoRI fragment containing the complete Bin1 coding region into pSRαMSV, a retroviral vector that lacks RNA packaging signals and includes a G418 resistance gene cassette. Briefly, recombinant virus was isolated from the media of

COS monkey cells cotransfected with pSRαMSV-Bin1 and pSVΨ-E-MLV, a proviral vector which provides the necessary retroviral packaging components. Recombinant virus were used to infect Rat1 fibroblasts and infected cell populations were selected by G418 selection. Expression of recombinant Bin1 in the Rat1 cell populations was confirmed by Northern and Western analysis.

Although the procedure above generated ecotropic Bin1 retroviruses limited to gene transfer to murine cells, those with skill in the art can easily generate amphotropic retroviruses that can transfer Bin1 to human cells. This is achieved by simply cotransfecting COS cells with pSRαMSV-Bin1 and pSVΨ-A-MLV, a packaging vector which encodes an amphotropic instead of ecotropic envelope glycoprotein [N. Landau et al, cited above]. Such vectors have been applied for use in gene therapies to attack human cancers.

D. Mammalian Expression Vectors

Bin1 mammalian cell expression vectors were constructed as follows and were used to generate the Bin1 proteins used in the following experiments. CMV-Bin1 was generated by subcloning a 1.6 kb EcoRI fragment from the full-length human Bin1 cDNA clone, p99f, that contained the entire predicted Bin1 coding sequence into pcDNA3 (Invitrogen), a mammalian cell expression vector that contains a cytomegalovirus enhancer/promoter and a 3' polyadenylation signal. CMV-HA-Bin1 was constructed by substituting a PvuII-EcoRI coding region fragment from CMV-Bin1 for an EcoRV-EcoRI fragment of neoCMV-hem rhoA, a RhoA expression plasmid that included an 8 residue N-terminal viral hemagglutinin (HA) epitope recognized by the monoclonal antibody 12CA5 [H. Niman et al, Proc. Natl. Acad. Sci. USA, 80:4949-4953 (1983)]. The HA-Bin1 polypeptide created included residues 1-47 from the N-

terminus of RhoA [Yeramian et al, Nucl. Acids Res., 15:1869 (1987)] and residues 52-451 of Bin1 [SEQ ID NO: 4]. This protein fusion added an N-terminal extension to Bin1 that allowed immunoprecipitation by anti-HA antibody 12CA5 [H. Niman et al, cited above]. CMV-Bin1ΔMBD deleted amino acid residues 270-377 [of SEQ ID NO: 4] in CMV-Bin1. It was constructed by ligating two separate PCR fragments generated by the 5' primer CCGGATCCGCGATGCTCTGGAACGTGGTGACG [nucleotides 60-80 of SEQ ID NO: 3] and the 3' primer GCGAATTCGTTGTCACTGTTC TTCTTTCTGCG (fragment encoding aa 1-269) [nucleotides 866-842, corresponding to the antisense strand of SEQ ID NO: 3] and the 5' primer CGGAATTCACCATGGGTTTCATGTTC AAGGTACAG [nucleotides 1191-1211 of SEQ ID NO: 3] and the 3' primer CCGCTCGAGTCATGGGACCCTCTCAGTGAAGTT (fragment encoding aa 378-451) [nucleotides 1415-1392, corresponding to the antisense strand of SEQ ID NO: 3]. This construction added the nonspecific amino acids EFTM at the fusion junction due to the restriction site added.

E. Immunoprecipitation

COS, MCF7, and IMR90 cells were cultured in Dulbecco's modified Eagle's media (DMEM) supplemented with 10% fetal bovine serum (Sigma) and 50 U/ml each penicillin and streptomycin (Fisher). Cells were transfected by a modified calcium phosphate protocol [C. Chen et al, Mol. Cell. Biol., 7:2745-2752 (1987)] and metabolically labeled 48 hr later. Rabbit antisera was raised to a GST fusion protein including amino acid residues 189-398 of Bin1 (GST-99Pst) [SEQ ID NO: 4], that included all of the MBD, using a commercial service (Rockland, Inc., Boyerstown, PA). Ten microliters of crude antisera or prebleed sera was used for immunoprecipitations from IMR90 or COS cells metabolically labeled 2-4 hr in DMEM media lacking methionine and cysteine (Gibco) with 75-125 μCi/ml

EXPRESS labeling reagent (NEN), washed with ice-cold phosphate-buffered saline, and extracted for 20 min on ice with RIPA buffer containing the protease inhibitors leupeptin, aprotinin, phenylmethylsulfonyl fluoride, and antipain [E. Harlow et al, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1988)]. Cell lysates were precleared by centrifugation at 20,000 g for 15 min at 4°C followed by 1 hr treatment with prebleed sera and 20 $\mu$l of a 1:1 slurry of protein G Sepharose beads at 4°C on a nutator (Pharmacia). Precleared lysates were immunoprecipitated 90 min at 4°C and then additional protein G beads were added and the incubation an additional 30 min. Beads were collected by brief centrifugation, washed four times with RIPA buffer, boiled in SDS gel loading buffer, fractionated on 10% gels, and fluorographed.

To establish that the Bin1 cDNA encoded a polypeptide similar to that found in normal cells, metabolically labeled extracts from IMR90 normal human diploid fibroblasts were subjected to immunoprecipitation. The results are described in Example 5 below.

Example 5 - Characterization of Bin1

A polyclonal antiserum was raised to a bacterially-expressed polypeptide derived from the unique central region of Bin1, in order to reduce the chance of crossreaction with Bin1-related proteins. When incubated with metabolically labeled extracts from COS cells transfected with CMV-Bin1, this antisera immunoprecipitated two polypeptides with apparent MW 70 kD and 45 kD. Each polypeptide was specifically recognized because their immunoprecipitation could be blocked by preincubating antisera with a molar excess of GST-Bin1 immunogen but not with unfused GST. In COS

cells transfected with CMV-HA-Bin1, only the 70 kD polypeptide was immunoprecipitated by an anti-HA monoclonal antibody. The IVT product from the full-length cDNA also had an apparent mobility of 70 kD. These data indicated that the 70 kD species was Bin1 and suggested that the 45 kD species was a Bin1-related polypeptide. Cells transfected with CMV-Bin1ΔMBD, a Bin1 deletion construct lacking the central Myc-binding domain (amino acid 270-377 of SEQ ID NO: 4), exhibited stable accumulation of a polypeptide whose predicted and apparent MW were both 42 kD. This result indicated that full-length Bin1 migrated aberrantly due to an MBD determinant at 70 kD in SDS polyacrylamide gels, instead of at the predicted MW of 50 kD. Only the 45 kD polypeptide was detected in untransfected MCF7 breast tumor cells, which lacked Bin1 RNA, or in cells transfected with empty vector. Thus, the 45 kD species was not a coprecipitant or an alternately processed or degraded form of Bin1. Consistent with its assignment as a Bin1-related protein, the 45 kD polypeptide could be detected by Western blotting.

Example 6 - Immunofluorescence Studies

~5 x $10^3$ HepG2 cells were seeded onto glass cover slips in 6 cm dishes and the next day transfected overnight with 4 $\mu$g CMV-Bin1 or pcDNA3. Two days later cells were washed and processed for immunofluorescence essentially as described [G. Prendergast et al, EMBO J., 10:757-766 (1991)], using 5 $\mu$g of protein A Sepharose-purified anti-Bin1 IgG and a 1:1000 dilution of fluorescein-conjugated anti-rabbit IgG (Cappel) as the secondary antibody. Stained cover slips were examined and analyzed on a Leitz confocal microscope.

In this manner, cell localization was examined by indirect cell immunofluorescence of transiently

transfected cultures of HepG2 hepatocarcinoma cells, which like MCF7 cells lack detectable Bin1 RNA (see below) and therefore provided an internal control for any crossreacting polypeptides. HepG2 cells transfected with CMV-Bin1 but not vector exhibited a speckled nuclear pattern of staining. The nuclear localization was consistent with the presence of a NLS in the primary sequence of Bin1 and with a nuclear site of interaction with Myc.

Example 7 - Inhibition of Myc Oncogenic Activity by Bin1

Since Bin1 was identified on the basis of its interaction with MB1, which is implicated in Myc transformation activity [J. Stone et al, Mol. Cell. Biol., 7:1697-1709 (1987) and B. Pulverer et al, Oncogene, 9:59-70 (1994)], the effects of Bin1 and the MBD deletion mutant Bin1ΔMBD (Example 4) were tested on cell transformation by Myc, adenovirus E1A, and SV40 T antigen in the Ras cooperation assay [H. Land et al, Nature, 304:596-602 (1983)] performed in primary rat embryo fibroblasts (REFs). Since the original clone #99 cDNA was partial and encoded essentially only the MBD, it was anticipated that the clone #99 ORF might act in a dominant negative manner to interfere with either endogenous Bin1. Therefore, the effects of a clone #99 expression vector (Example 4) on Myc transformation were also tested.

The ~0.5 kb murine cDNA [SEQ ID NO: 1] engineered with a 5' Kozak initiator methionine from pATG-99 was subcloned into pcDNA3 (a CMV enhancer/promoter vaccine; Invitrogen, San Diego, CA) to generate neoCMV-ATG99. REF culture and transfection was performed essentially as described [G. Prendergast et al, Genes Dev., 6:2429-2439 (1992)]. Briefly, secondary passage REFs seeded into 10 cm dishes were transfected overnight by a calcium

phosphate coprecipitation method [C. Chen et al, cited above] with 5 μg each of the oncogene plasmids and 10 μg of other plasmids indicated, then passaged 1:5 the next day and fed with normal growth media until foci were scored by methanol fixation and crystal violet staining 12-14 days later. In some experiments, 0.5 mg/ml G418 was added the day after passaging. The following oncogene plasmids were used in REF assays. LTR Hm, which contains a Moloney long terminal repeat-driven normal human c-myc gene, and pT22, which contains an activated H-ras gene, have been described [H. Land et al, cited above and A. Kelekar et al, Mol. Cell. Biol., 6:7-14 (1986)]. A nontransforming Myc frameshift mutant (MycFS) was constructed by digestion of LTR Hm with a unique Bst EII in exon 2 of the human c-myc gene, filling with Klenow enzyme, and self ligation. The MycFS polypeptide encoded by this mutant, LTR Hm/Bst, is frameshifted at amino acid residue 104, eliminating its biological function. This frameshift mutant was included to establish that the augmentation of foci formation by CMV-ATG99 was Myc-dependent. In some control experiments, NeoCMV T and p1A/neo, encoding SV40 T antigen and adenovirus E1A, respectively, were substituted for LTR Hm. Transformed foci were scored two weeks later.

The results of the REF focus formation experiments are shown in Fig. 3. On its own or with activated ras, Bin1 lacked transforming activity. However, when cotransfected with myc and ras, Bin1 selectively inhibited focus formation ~7-fold. Inhibition could be titered by decreasing the ratio of Bin1 to myc and ras vectors in the transfected DNA (data not shown). In contrast to the effect of full-length Bin1, but consistent with a dominant inhibitory effect, the murine vector neoCMV-ATG99 specifically augmented focus formation ~2- to 4-fold when cotransfected with myc and

ras. Bin1 also inhibited E1A-dependent transformation, consistent with the fact that E1A and Myc function similarly in biological assays [G. Evan et al, Cell, 69:119-128 (1992); H. Land et al, cited above; H. Ruley, Nature, 304:602-606 (1983); and L. Rao et al, Proc. Natl. Acad. Sci. USA, 89:7742-7746 (1992)]. However, Bin1 did not affect T antigen-dependent transformation. This result indicated that the inhibition of Myc and E1A was not due to toxicity or nonspecific inhibition of the transformed phenotype. Notably, Bin1ΔMBD significantly inhibited E1A but not Myc. The lack of an effect of Bin1ΔMBD on Myc transformation could not be explained by protein instability, because Bin1ΔMBD had been shown to stably accumulate in transfected COS cells and could inhibit E1A transformation. Although the means by which Bin1 and Bin1ΔMBD inhibited E1A was unclear, an important implication of this result was that Bin1 inhibited E1A and Myc by different mechanisms, an interpretation consistent with the differential binding of these oncoproteins to the MBD represented in GST-99. Supporting the notion that Bin1 was incompatible with Myc or E1A transformation, exogenous Bin1 message accumulated in REF cell populations derived from transformation with T antigen but not with Myc or E1A; in contrast, Bin1ΔMBD message accumulated in REFs transformed by either Myc or E1A. There is a possibility that a reduced activity of Bin1ΔMBD revealed intrinsic differences in the sensitivity of E1A and Myc to Bin1 inhibition. However, with this caveat, it was concluded that Bin1 physiologically interacted with and inhibited Myc, since deletion of a Bin1 domain sufficient for association in vitro was necessary for its inhibition activity in vivo.

Example 8 - Northern Analysis

Northern analysis was performed to examine Bin1 expression patterns. Total cytoplasmic RNA from tumor cell lines was isolated [G. Prendergast et al, Mol. Cell. Biol., 9:124-134 (1989)] and hybridized to [$^{32}$P]-labeled Bin1 cDNA probes [G. Church et al, Proc. Natl. Acad. Sci. USA, 81:1991-1995 (1984)].

Ubiquitous expression in normal murine and human cells was observed. In the mouse, RNA levels were highest in embryo, adult brain, and adult muscle but lower levels were seen in all other tissues examined. In embryo and brain, at least two transcripts could be resolved, suggesting alternate splicing or differential usage of initiation or polyadenylation sites in some cells. In human cells, RNA levels were similar in WI-38 normal diploid fibroblasts and tumor cells derived from several different tissues. However, Bin1 message levels were undetectable in HepG2 hepatocarcinoma and MCF7 breast carcinoma cells and were >10-fold reduced in SK-CO-1 colon carcinoma cells. Further examination revealed similar deficits in 5/6 breast and 3/6 cervix carcinomas, and in 4/7 liver and 1/2 lung carcinomas. In total, loss of Bin1 message was observed in 14/27 carcinoma cell lines examined.

Example 9 - Colony Formation Assays

The functional significance of deficits in Bin1 message levels in certain tumor cells (as in Example 8) was suggested by G418-resistant cell colony formation experiments performed in four cell lines available from the American Type Culture Collection (Rockville, Maryland) that either contained (HeLa) or lacked (HepG2, MCF7, SAOS-2) endogenous Bin1 RNA.

Colony formation assays were performed in the following manner. $3 \times 10^5$ cells in 3 cm dishes were

transfected overnight with 2 µg CMV-Bin1 (described in Example 4) or an empty vector, using Lipofectamine (Gibco/BRL). Cells were passaged 48 hr after transfection at a 1:10 ratio into 6 cm dishes containing media with ~0.6 mg/ml G418, which permits selection for the neomycin gene present on each plasmid. Drug-resistant cell colonies were scored by crystal violet staining 2-3 weeks later. At least three trials for each cell line were performed and colonies were scored in triplicate dishes.

HepG2, MCF7, and SAOS-2 cells transfected with a Bin1 vector exhibited approximately 3-fold fewer colonies relative to cells transfected with empty vector, whereas no significant difference in HeLa colony formation was seen (Fig. 4). Cell populations derived from pooled colonies which emerged from Bin1-transfected HepG2 cultures showed no evidence of expression, when examined by immunoprecipitation, consistent with an incompatibility with cell growth. From this data, it can be concluded that the RNA deficits seen in carcinoma cells are functionally significant and that Bin1 can inhibit tumor cell growth, consistent with a tumor suppressor function.

<u>Example 10 - Rearrangement and loss of expression of the Bin1 gene in liver and breast cancer cells</u>

Because Bin1 had been demonstrated to inhibit Myc-dependent cell transformation and tumor cell growth, the following study was performed to determine if the Bin1 gene is mutated in human tumor cells. The initial experiment was to perform Southern analysis of the genomic DNA from a panel of human tumor cell lines including HeLa [cervix, ATCC CCL 2], SK-CO-1 [colon, ATCC HTB 39], HT-29 [colon, ATCC HTB 38], DU145 [prostate, ATCC HTB 41], PC-3 [prostate, ATCC CRL 1435], LNCaP

[prostate, ATCC CRL 1740]; T24 [bladder, ATCC HTB4]; MCF7 [breast, ATCC HTB 22]; HepG2 [liver, ATCC HB 8065]; Rh-30 [myosarcoma, E.C. Douglass et al, "A specific chromosomal abnormality in rhabdosarcoma, Cytogenet. Cell Genet., 45:148-155 (1987)]; Raji [lymphoma, ATCC CCL 86]. DNA from WI-38 normal diploid fibroblasts [ATCC CCL 75] was used as a source of normal DNA.

DNAs were isolated by standard methods (Sambrook et al, cited above) and 5 $\mu$g per sample was treated with HindIII restriction endonuclease. Restricted DNA was fractionated on a 0.65% agarose gel which was denatured 2 x 15 minutes in 1.5 M NaCl/0.5M NaOH, neutralized 2 x 30 minutes in 1.5 M NaCl/0.5 TrisCl pH 8, and then blotted to a charged nylon membrane (Stratagene, La Jolla CA). The blot was crosslinked by UV irradiation and hybridized in a commercial hybridization solution with a random-primed $^{32}$P-labeled Bin1 cDNA probe according to the vendor's instructions (Amersham, Cambridge UK). The blot was washed 1 x 10 minutes with 2X SSC/0.1% SDS at 20°C and then 2 x 10 minutes with 0.2 SSC/0.1% SDS at 65°C before being exposed to X-ray film (DuPont, Wilmington DE).

Two bands of >20 kb and 6.5 kb were observed in all the genomic DNAs except for HepG2, a liver carcinoma cell line, where an additional band of ~3.5 kb was seen. Following this observation, a second Southern analysis was performed on a panel of 9 liver carcinoma cell lines, including Huh1, Huh2, HepG2 [ATCC HB8065], Hep3B [ATCC HB8064], Hep43, Hep63, HLF [ATCC CCL 199], NCH2, and NHep40 (provided by Dr. D. Simon, Medical College of Pennsylvania). Conditions were the same as above except that PstI restriction endonuclease was used.

Five bands of 2.5, 1.8, 1.5, 0.95, and 0.75 kb were observed in WI-38 normal DNA. Four of the nine liver tumor DNAs (HepG2, Hep3B, NCH2, and NHep40) exhibited an

additional band of 2.9-3.3 kb. These data corroborated the previous results and indicated that Bin1 may be mutated during the development of human hepatocarcinoma.

Northern analysis on RNA isolated from the initial panel of human tumor cells, including HepG2 liver carcinoma cells, was performed to examine Bin1 expression. A similar analysis of RNAs isolated from mouse embryo or adult tissues was also performed. Total cytoplasmic RNA was purified by standard methods (Sambrook et al, cited above) and 15 μg was fractionated on a 1% formaldehyde agarose gel and blotted as described [G.C. Prendergast and M.D. Cole, Mol. Cell. Biol., 9: 124-134 (1989)]. A commercial Northern blot containing RNA from normal human brain, heart, kidney, lung, liver, skeletal muscle, pancreas, and placenta (Clontech, Palo Alto CA) was also analyzed. Using the same procedure and conditions as above, the Northern blots were hybridized with Bin1 cDNA probe, washed, and exposed to X-ray film.

A Bin1-specific ~2.2 kb RNA was observed in all tissues and cell lines except HepG2 and the breast carcinoma cell line MCF7. This result indicated that Bin1 was ubiquitously expressed and that mutation in HepG2 cells was correlated with loss of expression. This result directly supports the utility of RT-PCR and FISH for diagnosing Bin1 loss in liver and breast cancer biopsies.

All documents cited above are incorporated herein by reference. Numerous modifications and variations of the present invention are included in the above-identified specification and are expected to be obvious to one of skill in the art. Such modifications and alterations to the compositions and processes of the present invention are believed to be encompassed in the scope of the claims appended hereto.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Wistar Institute of Anatomy & Biology

(ii) TITLE OF INVENTION: Murine and Human Box-Dependent Myc-Interacting Protein (BIN1) and Uses Therefor

(iii) NUMBER OF SEQUENCES: 7

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Howson and Howson
(B) STREET: Spring House Corporate Cntr, P O Box 457
(C) CITY: Spring House
(D) STATE: Pennsylvania
(E) COUNTRY: USA
(F) ZIP: 19477

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: WO
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/435,454
(B) FILING DATE: 05-MAY-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Bak, Mary E.
(B) REGISTRATION NUMBER: 31,215
(C) REFERENCE/DOCKET NUMBER: WST60APCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 215-540-9200
(B) TELEFAX: 215-540-5818

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 402 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..399

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GAG ATC AGA GTG AAC CAT GAG CCA GAG CCG GCC AGT GGG GCC        42
Glu Ile Arg Val Asn His Glu Pro Glu Pro Ala Ser Gly Ala
  1               5                  10

TCA CCC GGG GCT GCC ATC CCC AAG TCC CCA TCT CAG CCA GCA        84
Ser Pro Gly Ala Ala Ile Pro Lys Ser Pro Ser Gln Pro Ala
 15                  20                  25

GAG GCC TCC GAG GTG GTG GGT GGA GCC CAG GAG CCA GGG GAG       126
Glu Ala Ser Glu Val Val Gly Gly Ala Gln Glu Pro Gly Glu
     30                  35                  40

ACA GCA GCC AGT GAA GCA ACC TCC AGC TCT CTT CCG GCT GTG       168
Thr Ala Ala Ser Glu Ala Thr Ser Ser Ser Leu Pro Ala Val
         45                  50                  55

GTG GTG GAG ACC TTC TCC GCA ACT GTG AAT GGG GCG GTG GAG       210
Val Val Glu Thr Phe Ser Ala Thr Val Asn Gly Ala Val Glu
             60                  65                  70

GGC AGC GCT GGG ACT GGA CGC TTG GAC CTG CCC CCG GGA TTC       252
Gly Ser Ala Gly Thr Gly Arg Leu Asp Leu Pro Pro Gly Phe
                 75                  80

ATG TTC AAG GTT CAA GCC CAG CAT GAT TAC ACG GCC ACT GAC       294
Met Phe Lys Val Gln Ala Gln His Asp Tyr Thr Ala Thr Asp
 85                  90                  95

ACT GAT GAG CTG CAA CTC AAA GCT GGC GAT GTG GTG TTG GTG       336
Thr Asp Glu Leu Gln Leu Lys Ala Gly Asp Val Val Leu Val
    100                 105                 110

ATT CCT TTC CAG AAC CCA GAG GAG CAG GAT GAA GGC TGG CTC       378
Ile Pro Phe Gln Asn Pro Glu Glu Gln Asp Glu Gly Trp Leu
        115                 120                 125

ATG GGT GTG AAG GAG AGC GAC TGA                               402
Met Gly Val Lys Glu Ser Asp
            130
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 133 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Glu | Ile | Arg | Val | Asn | His | Glu | Pro | Glu | Pro | Ala | Ser | Gly | Ala | Ser |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Pro | Gly | Ala | Ala | Ile | Pro | Lys | Ser | Pro | Ser | Gln | Pro | Ala | Glu | Ala |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Ser | Glu | Val | Val | Gly | Gly | Ala | Gln | Glu | Pro | Gly | Glu | Thr | Ala | Ala |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Ser | Glu | Ala | Thr | Ser | Ser | Ser | Leu | Pro | Ala | Val | Val | Val | Glu | Thr |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Phe | Ser | Ala | Thr | Val | Asn | Gly | Ala | Val | Glu | Gly | Ser | Ala | Gly | Thr |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Gly | Arg | Leu | Asp | Leu | Pro | Pro | Gly | Phe | Met | Phe | Lys | Val | Gln | Ala |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Gln | His | Asp | Tyr | Thr | Ala | Thr | Asp | Thr | Asp | Glu | Leu | Gln | Leu | Lys |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Ala | Gly | Asp | Val | Val | Leu | Val | Ile | Pro | Phe | Gln | Asn | Pro | Glu | Glu |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Gln | Asp | Glu | Gly | Trp | Leu | Met | Gly | Val | Lys | Glu | Ser | Asp | | |
| | | | | 125 | | | | | 130 | | | | | |

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1925 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 60..1412

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
GAATTCCGTG CTGGTTGAGC TTGCTCATCT CCTTGTGGAA GTTTTCCTCC          50

AGGCCCGCG ATG CTC TGG AAC GTG GTG ACG GCG GGA AAG ATC           92
          Met Leu Trp Asn Val Val Thr Ala Gly Lys Ile
           1               5                  10

GCC AGC AAC GTG CAG AAG AAG CTC ACC CGC GCG CAG GAG AAG        134
Ala Ser Asn Val Gln Lys Lys Leu Thr Arg Ala Gln Glu Lys
             15                  20                  25

GTT CTC CAG AAG CTG GGG AAG GCA GAT GAG ACC AAG GAT GAG        176
Val Leu Gln Lys Leu Gly Lys Ala Asp Glu Thr Lys Asp Glu
                 30                  35

CAG TTT GAG CAG TGC GTC CAG AAT TTC AAC AAG CAG CTG ACG        218
Gln Phe Glu Gln Cys Val Gln Asn Phe Asn Lys Gln Leu Thr
 40                  45                  50

GAG GGC ACC CGG CTG CAG AAG GAT CTC CGG ACC TAC CTG GCC        260
Glu Gly Thr Arg Leu Gln Lys Asp Leu Arg Thr Tyr Leu Ala
     55                  60                  65

TCC GTC AAA GCC ATG CAC GAG GCT TCC AAG AAG CTG AAT GAG        302
Ser Val Lys Ala Met His Glu Ala Ser Lys Lys Leu Asn Glu
         70                  75                  80

TGT CTG CAG GAG GTG TAT GAG CCC GAT TGG CCC GGC AGG GAT        344
Cys Leu Gln Glu Val Tyr Glu Pro Asp Trp Pro Gly Arg Asp
             85                  90                  95

GAG GCA AAC AAG ATC GCA GAG AAC AAC GAC CTG CTG TGG ATG        386
Glu Ala Asn Lys Ile Ala Glu Asn Asn Asp Leu Leu Trp Met
                100                 105

GAT TAC CAC CAG AAG CTG GTG GAC CAG GCG CTG CTG ACC ATG        428
Asp Tyr His Gln Lys Leu Val Asp Gln Ala Leu Leu Thr Met
110                 115                 120

GAC ACG TAC CTG GGC CAG TTC CCC GAC ATC AAG TCA CGC ATT        470
Asp Thr Tyr Leu Gly Gln Phe Pro Asp Ile Lys Ser Arg Ile
    125                 130                 135

GCC AAG CGG GGG CGC AAG CTG GTG GAC TAC GAC AGT GCC CGG        512
Ala Lys Arg Gly Arg Lys Leu Val Asp Tyr Asp Ser Ala Arg
        140                 145                 150
```

```
CAC CAC TAC GAG TCC CTT CAA ACT GCC AAA AAG AAG GAT GAA      554
His His Tyr Glu Ser Leu Gln Thr Ala Lys Lys Lys Asp Glu
            155                 160                 165

GCC AAA ATT GCC AAG GCC GAG GAG GAG CTC ATC AAA GCC CAG      596
Ala Lys Ile Ala Lys Ala Glu Glu Glu Leu Ile Lys Ala Gln
                170                 175

AAG GTG TTT GAG GAG ATG AAT GTG GAT CTG CAG GAG GAG CTG      638
Lys Val Phe Glu Glu Met Asn Val Asp Leu Gln Glu Glu Leu
180                 185                 190

CCG TCC CTG TGG AAC AGC CGC GTA GGT TTC TAC GTC AAC ACG      680
Pro Ser Leu Trp Asn Ser Arg Val Gly Phe Tyr Val Asn Thr
    195                 200                 205

TTC CAG AGC ATC GCG GGC CTG GAG GAA AAC TTC CAC AAG GAG      722
Phe Gln Ser Ile Ala Gly Leu Glu Glu Asn Phe His Lys Glu
        210                 215                 220

ATG AGC AAG CTC AAC CAG AAC CTC AAT GAT GTG CTG GTC GGC      764
Met Ser Lys Leu Asn Gln Asn Leu Asn Asp Val Leu Val Gly
            225                 230                 235

CTG GAG AAG CAA CAC GGG AGC AAC ACC TTC ACG GTC AAG GCC      806
Leu Glu Lys Gln His Gly Ser Asn Thr Phe Thr Val Lys Ala
                240                 245

CAG CCC AGA AAG AAA AGT AAA CTG TTT TCG CGG CTG CGC AGA      848
Gln Pro Arg Lys Lys Ser Lys Leu Phe Ser Arg Leu Arg Arg
250                 255                 260

AAG AAG AAC AGT GAC AAC GCG CCT GCA AAA GGG AAC AAG AGC      890
Lys Lys Asn Ser Asp Asn Ala Pro Ala Lys Gly Asn Lys Ser
    265                 270                 275

CCT TCG CCT CCA GAT GGC TCC CCT GCC GCC ACC CCC GAG ATC      932
Pro Ser Pro Pro Asp Gly Ser Pro Ala Ala Thr Pro Glu Ile
        280                 285                 290

AGA GTC AAC CAC GAG CCA GAG CCG GCC GGC GGG GCC ACG CCC      974
Arg Val Asn His Glu Pro Glu Pro Ala Gly Gly Ala Thr Pro
            295                 300                 305

GGG GCC ACC CTC CCC AAG TCC CCA TCT CAG CCA GCA GAG GCC     1016
Gly Ala Thr Leu Pro Lys Ser Pro Ser Gln Pro Ala Glu Ala
                310                 315

TCG GAG GTG GCG GGT GGG ACC CAA CCT GCG GCT GGA GCC CAG     1058
Ser Glu Val Ala Gly Gly Thr Gln Pro Ala Ala Gly Ala Gln
320                 325                 330
```

GAG CCA GGG GAG ACT TCT GCA AGT GAA GCA GCC TCC AGC TCT 1100
Glu Pro Gly Glu Thr Ser Ala Ser Glu Ala Ala Ser Ser Ser
335 340 345

CTT CCT GCT GTC GTG GTG GAG ACC TTC CCA GCA ACT GTG AAT 1142
Leu Pro Ala Val Val Val Glu Thr Phe Pro Ala Thr Val Asn
350 355 360

GGC ACC GTG GAG GGC GGC AGT GGG GCC GGG CGC TTG GAC CTG 1184
Gly Thr Val Glu Gly Gly Ser Gly Ala Gly Arg Leu Asp Leu
365 370 375

CCC CCA GGT TTC ATG TTC AAG GTA CAG GCC CAG CAC GAC TAC 1226
Pro Pro Gly Phe Met Phe Lys Val Gln Ala Gln His Asp Tyr
380 385

ACG GCC ACT GAC ACA GAC GAG CTG CAG CTC AAG GCT GGT GAT 1268
Thr Ala Thr Asp Thr Asp Glu Leu Gln Leu Lys Ala Gly Asp
390 395 400

GTG GTG CTG GTG ATC CCC TTC CAG AAC CCT GAA GAG CAG GAT 1310
Val Val Leu Val Ile Pro Phe Gln Asn Pro Glu Glu Gln Asp
405 410 415

GAA GGC TGG CTC ATG GGC GTG AAG GAG AGC GAC TGG AAC CAG 1352
Glu Gly Trp Leu Met Gly Val Lys Glu Ser Asp Trp Asn Gln
420 425 430

CAC AAG AAG CTG GAG AAG TGC CGT GGC GTC TTC CCC GAG AAC 1394
His Lys Lys Leu Glu Lys Cys Arg Gly Val Phe Pro Glu Asn
435 440 445

TTC ACT GAG AGG GTC CCA TGACGGCGGG GCCCAGGCAG CCTCCGGGCG 1442
Phe Thr Glu Arg Val Pro
450

TGTGAAGAAC ACCTCCTCCC GAAAAATGTG TGGTTCTTTT TTTTGTTTTG 1492

TTTTCGTTTT TCATCTTTTG AAGAGCAAAG GGAAATCAAG AGGAGACCCC 1542

CAGGCAGAGG GGCGTTCTCC CAAAGTTTAG GTCGTTTTCC AAAGAGCCGC 1592

GTCCCGGCAA GTCCGGCGGA ATTCACCAGT GTTCCTGAAG CTGCTGTGTC 1642

CTCTAGTTGA GTTTCTGGCG CCCCTGCCTG TGCCCGCATG TGTGCCTGGC 1692

CGCAGGGCGG GGCTGGGGGC TGCCGAGCCA CCATACTTAA CTGAAGCTTC 1742

GGCCGCACCA CCCGGGGAAG GGTCCTCTTT TCCTGGCAGC TGCTGTGGGT 1792

```
GGGGCCCAGA CACCAGCCTA GCCTGCTCTG CCCCGCAGAC GGTCTGTGTG    1842

CTGTTTGAAA ATAAATCTTA GTGTTCAAAA CAAAATGAAA CAAAAAAAAA    1892

AATGATAAAA ACTCTCAAAA AAACAAGGAA TTC                      1925
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 451 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Leu Trp Asn Val Val Thr Ala Gly Lys Ile Ala Ser Asn Val
 1               5                  10                  15

Gln Lys Lys Leu Thr Arg Ala Gln Glu Lys Val Leu Gln Lys Leu
                20                  25                  30

Gly Lys Ala Asp Glu Thr Lys Asp Glu Gln Phe Glu Gln Cys Val
                35                  40                  45

Gln Asn Phe Asn Lys Gln Leu Thr Glu Gly Thr Arg Leu Gln Lys
                50                  55                  60

Asp Leu Arg Thr Tyr Leu Ala Ser Val Lys Ala Met His Glu Ala
                65                  70                  75

Ser Lys Lys Leu Asn Glu Cys Leu Gln Glu Val Tyr Glu Pro Asp
                80                  85                  90

Trp Pro Gly Arg Asp Glu Ala Asn Lys Ile Ala Glu Asn Asn Asp
                95                 100                 105

Leu Leu Trp Met Asp Tyr His Gln Lys Leu Val Asp Gln Ala Leu
               110                 115                 120

Leu Thr Met Asp Thr Tyr Leu Gly Gln Phe Pro Asp Ile Lys Ser
               125                 130                 135

Arg Ile Ala Lys Arg Gly Arg Lys Leu Val Asp Tyr Asp Ser Ala
               140                 145                 150

Arg His His Tyr Glu Ser Leu Gln Thr Ala Lys Lys Lys Asp Glu
               155                 160                 165
```

Ala Lys Ile Ala Lys Ala Glu Glu Glu Leu Ile Lys Ala Gln Lys
170 175 180

Val Phe Glu Glu Met Asn Val Asp Leu Gln Glu Glu Leu Pro Ser
185 190 195

Leu Trp Asn Ser Arg Val Gly Phe Tyr Val Asn Thr Phe Gln Ser
200 205 210

Ile Ala Gly Leu Glu Glu Asn Phe His Lys Glu Met Ser Lys Leu
215 220 225

Asn Gln Asn Leu Asn Asp Val Leu Val Gly Leu Glu Lys Gln His
230 235 240

Gly Ser Asn Thr Phe Thr Val Lys Ala Gln Pro Arg Lys Lys Ser
245 250 255

Lys Leu Phe Ser Arg Leu Arg Arg Lys Lys Asn Ser Asp Asn Ala
260 265 270

Pro Ala Lys Gly Asn Lys Ser Pro Ser Pro Pro Asp Gly Ser Pro
275 280 285

Ala Ala Thr Pro Glu Ile Arg Val Asn His Glu Pro Glu Pro Ala
290 295 300

Gly Gly Ala Thr Pro Gly Ala Thr Leu Pro Lys Ser Pro Ser Gln
305 310 315

Pro Ala Glu Ala Ser Glu Val Ala Gly Gly Thr Gln Pro Ala Ala
320 325 330

Gly Ala Gln Glu Pro Gly Glu Thr Ser Ala Ser Glu Ala Ala Ser
335 340 345

Ser Ser Leu Pro Ala Val Val Val Glu Thr Phe Pro Ala Thr Val
350 355 360

Asn Gly Thr Val Glu Gly Gly Ser Gly Ala Gly Arg Leu Asp Leu
365 370 375

Pro Pro Gly Phe Met Phe Lys Val Gln Ala Gln His Asp Tyr Thr
380 385 390

Ala Thr Asp Thr Asp Glu Leu Gln Leu Lys Ala Gly Asp Val Val
395 400 405

Leu Val Ile Pro Phe Gln Asn Pro Glu Glu Gln Asp Glu Gly Trp
410 415 420

```
Leu Met Gly Val Lys Glu Ser Asp Trp Asn Gln His Lys Lys Leu
                425                 430                 435

Glu Lys Cys Arg Gly Val Phe Pro Glu Asn Phe Thr Glu Arg Val
                440                 445                 450

Pro
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Glu Asp Ile Trp Lys Lys Phe Glu Leu Leu Pro Thr Pro Pro Leu
  1               5                  10                  15

Ser
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14985 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1332
(D) OTHER INFORMATION: /note= "unsequenced segment"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 3225
(D) OTHER INFORMATION: /note= "unsequenced segment"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 7209
(D) OTHER INFORMATION: /note= "unsequenced segment"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 11097
(D) OTHER INFORMATION: /note= "unsequenced segment"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..324
(D) OTHER INFORMATION: /note= "Exon 1"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 325..1618
(D) OTHER INFORMATION: /note= "Exon 2"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1619..3174
(D) OTHER INFORMATION: /note= "Exon 3"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 3175..4365
(D) OTHER INFORMATION: /note= "Exon 4"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 4441..11518
(D) OTHER INFORMATION: /note= "Exon 5"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 11519..11850
(D) OTHER INFORMATION: /note= "Exon 6"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 11851..12240
(D) OTHER INFORMATION: /note= "Exon 7"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 12241..14129
(D) OTHER INFORMATION: /note= "Exon 8"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 14130..14985
(D) OTHER INFORMATION: /note= "Exon 9"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TCGATCTAGC AGGATGAGCT GCCGTCCCTG TGGAACAGGT GAGGCCCGGC 50
ACGGTGCCCA GCCTGCGTGG GGCAGTGTCC AGTCTGCGTG CTGCAGTGCC 100
CAGTCTGCGT GCTGTGGTGC CCAGCCTGCG TGCTGCGGTG TCCAGTCTGC 150
GTGCTGCGGT GCCCAGCCTG TGTGCTGCAG TGTCCAGCCT GAGTGCTGCG 200
GTGCCTGGGC CCTCTCTGGT TTGTGCCTCT GATGAGCGTG TGTGGTCGCT 250
CGTGGGTGGG TATTTCTGAG TTGCTGTCCT GACCTGCCTG TTCACCTGGC 300
CCCCATCCTT CCGCCCTTCT GTAGCCGCGT AGGTTTCTAC GTCAACACGT 350
TCCAGAGCAT CGCGGGCCTG GAGGAAAACT TCCACAAGGA GATGAGCAAG 400
GTAGGCCATG GGGACCCCTC TGAGGGGCCA CACCCCACCC TGGCCGAGGG 450
TCAGAGTCAG AATCGTGGGA GGGGCAGCCT GAACTCCTCC TTCCCTGCCA 500
GGTTCAGCAC ACACCGGTGA CCACAGGGCT CCCTCCCGGC CCTGGTGGAA 550
CAGCCCCCTT CAGGAGTGCC TGGGCCCCAG GAAGGGCACC AGGGCATGCT 600
GGGGAGGCTT TGAGAGTGTC CTGGGTCCTT GCCTGGGTAG ACCACCTGAG 650
AATGTAGACC AGGCCCTCTC AAACTGTGAA TGTGTTCTCT AGCAACCTTG 700
GCCCAGGGAG TGCAGTGTTA GACAATGGTG GCAGTTTCTC AATTTGTAGG 750
AAAAAATTAC TAGCATTTTC TCAATTTTAA TTTTTCTCCC ATGCTTTTGA 800
GACATTTAAA TAGGCCTGTT TCGTCTGCGT GAATCCACTG TGCGACAGGG 850
TCTGGTCTGA TTAGCGTACC TTTCTCTCTT GTGTTATTTT CATTTTAACC 900
TATCTGGCAA CTAAAACGCA AAGCTGTTGA ACTTTGCAGT TGGAGAGACC 950
CAGGGCTGGC ACCCCCGCCC CCAGTGGTGG GTCTGACTTT GTCTCTTGGG 1000
CCCAGATGGA TGAAGTGATG ATGGTGGATT CCAGGCCCAG GGTCGGAGGA 1050
GGAAAAGCAG CTTGGGGCCT CCCCTGACTC ACACTTCTAA GTACGGTTTC 1100
CTGCCTTTTA GACTTTCTCT TTCCTTCTTA ACTTTTCCTT TTTCCCTTCG 1150
AAGATTGGAG ACAACTTACG AAAAGTTTAA AAAATAGAGA AAGGTGTTTC 1200
AGAGAAGGAA CATTTATATA AAATTAGTCT GTAAAAATAT ATGCCACAGA 1250

ACATAGTTGT TTATAAATAG GATGAAGATT CGGCCTGAGC TCCTAGTGGC 1300

CAAATCAAAG AAANGAATAT CTCAGTAATG ANGGNGTTNA TAATGGGGCA 1350

TTTTAGGATA TCCCTAATAA GTAGGGAGGG GAAAAGATGG TTCCAAGAGA 1400

CAGCAGGCAT TAGNGATATA GCCCANTTGC CCGGTTTTNG NNGCCCCCTN 1450

GGTTTTTGGG GGGNTTGGTG TGCNAGTTCC TTGCTCTCAC AGATGGGGAA 1500

AGAGGAGCTT GTTCCTGGCA GGGGCTGGGG GTGGTGGGGA GAAGCAGAGG 1550

TGTTTGGGGA AGGTGGGGCC GTTTGGTGGC CTTGGAGGCC CCCCACCTCC 1600

TCACTGTCTC TCCTGCAGCT CAACCAGAAC CTCAATGATG TGCTGGTCGG 1650

CCTGGAGAAG CAACACGGGA GCAACACCTT CACGGTCAAG GCCCAGCCCA 1700

GGTGCGTGCG GGGAGAGCCC TGGCGCCCCT GACTGTGTGC ACGGCAGGGG 1750

CAGGGCTCCT TCCTGTGACC CTGTTGGTGC CCTCCCCTGG TCCCCCATGG 1800

GTTTGGCCTT GGGGGTCTAG GGACCTTCCT GTCTTGGCCT CTCTGTGCTC 1850

AGGGAGGCAG GTGAGGGCAG GTCTCTGTCT CGAATGTCCC TGCCCCTCTG 1900

GCTGTGTTCG TCGAGGAAGG AGCACTCTGG GGAGTCCGCG GGTACCCTGA 1950

GCCGGCTGAC CCCCTCATTG TGGAGCACGA GCATCCAGGG TTGGGGTGGG 2000

CAGCCTGCTC AGCTTTGGGG ACTGGGGGGT GTGAACAGGA CTGAAAGACT 2050

CCGGGGTGTG CAGTCCTCTC AGAGCAGGGA GATAGCACCG CCCTTCCTCT 2100

CCTGCTNGTG GNAAAAGATC ATGTCCCTGG ATGGCAGCAT TGTGCTCAAC 2150

CACANGAGCA TCCTCTTCCT GTCCTCAGCC TCAGCCCCTC CGGGAATCCC 2200

AGCTGCAAGG AGGCCTCTGT TTCCTGAGGG GAAACCATGA GGGAGGAGGG 2250

AAATGCCTTG CTTTCCTGGC TGTGGATCAG AGGAAGCAGC GAGCCTGGGA 2300

CTTCCCCTCC CTTNTGGCCA TGTGTGCATG TGTGTGTGTG AGGGGGACTG 2350

TGTGTGACAG GTGTAAGTGT GTGCATACCC ACACACATAT CACAGCAGAA 2400

CGCAGAGAAC ACCGATGGAC TCTGTAAAAC AGGGCGACTG TCTGCTTCTT 2450

GGGGTATTGC CTGGGATGAT GAGGGTATCG GGTGGTGGTG ATTGCCCCCT 2500

CCTTCCCTGA ACATAAAATA GTTGTGGCTG AGAGAGGGGC CATGGTGACC 2550

TGAGGCTGGG AGTGGGGAGG TTAGGACGGT GGCGTTGTGG TGGTGGTTGG 2600
GGGGGTGGGT AGGGGGGTGG GGGTTGGGAT AAAGCCAAAA GGTGAATTCA 2650
AGGTCGGGCA GGGAGGGACA GCTGCCTGGC CTGTAGGCAC AGGTGGGAAC 2700
ANTGGGATGG ATCAGCAGGG GGTAAGTGGG GCCGTCCTGG CCAGAACCAT 2750
GGCTCCCCTC AGGAAGGAGG TGGAGGGAAG AGAGAGGGGC AGTAGAGGCC 2800
CAGGAGTCTC CCTTCCAGCA GAGAGGCCTC TTGTGCACTN TGTGCTCGCC 2850
TGGGGGCCTT TTCTGGCACT NTGGGCACAC CTGGAGCTCC TGGGGACTGG 2900
GACCACAGGC AGGGTGACTA TCCACTGCCC CGAGCCTCCT GCCCCTCACC 2950
AGGCCCTGTT AGCATCACCT CGGGCACCTG GCCACAGCAG GGGCCAGTCA 3000
GGGCACCCCG GGATAGCACG CCCAGGCCCT GTGCAAGGCC TCTGGCACTT 3050
AGGAGAGGCT TTTGCCCCTT TGTCCTCTGA GCAGAAGGGT TGGCAAAGAG 3100
GGAAGGGGAC AGGCCAGTTC TGCACCTGGC CTTTCTCCAG AATGAAGGCC 3150
TCCACCTCCC GTCCGTCCCC ACAGAAAGAA AAGTAAACTG TTTTCGCGGC 3200
TGCGCAGAAA GAAGAACAGG TACCNAGGAT GGTGCCCCGG GGCCAACCCA 3250
GGGGCAATTT AAGCAAATGG AGGAAGGNGG GNTGGGGAAA GGAGGCNTGG 3300
GGGAGGGCCA GGGAGGGTGA NAATGCAGCA AATGTGGGGG TTTTTTGTTT 3350
TTTAAAACAA ATTGTATGTG TATACCATAT ACTTATACAT ATATTCTTTA 3400
AGGAGAATAC ATTCCCATAA AACACAAATT CCAGAAGGAA AGATGGTGTC 3450
AGCGACATCT CTTACGNTGT TCCACTGTTT GCCCTCAGGT GANTCGGTCA 3500
CTGGTTCCTG CTGGATGNTT GTAGATGTGC ACTGTCCAGC ACAGGAGCCA 3550
GTTACCCCAT GGGGCTATTG AGCACTTGAA ACTGGCCAGT GTGACCGGGC 3600
AGCGGAACTT TTCATTTGAA TTACACATAA ATTTCATTGC TTTGAGTTTG 3650
CATTGCCGCC TGTGGCTAGT GGCAACCGTA CTGGGCAGCA CTTTTCTAGG 3700
CGTCTCTGTG CAGGTTCTGG TAGAGAATTT TCTCCCTGCA CCTTCGCCCC 3750
TGTGCCTGGG GTGCACAGCA TCACACCACC TCCGCCTTGG GTTCTGGCAC 3800
TGAACGCCAT GGCTCAGGAC CTGTCCCCTC CATCGCCAGC TGCCCACTCC 3850

TCTGTGATGA GGACGCCTCT CTTAGTTTGT CCAGGCCCTG CTTGTGGCCT 3900

CCAGCAGCCG AGAGGACAGG AGAGCCCAAG GTCTAGAGAC ATGTACCAGG 3950

GTGCTGTGAT GGACAGGCAG GGAGGGCAGC AGGCTGGGGA GCAGACCCCA 4000

GAACAGAGGG GCTGCTGCGT GTGGTGTGGG AGACTCACTG TGCCTCTAGG 4050

ATGTCTGGCT TTCTCCTGCT GTGGATCTTG GGCTGTCAGC ATGGGCCCTG 4100

GTGGACCCCA TGGAGCCTGT GGGGTGGTTG GTCTGGTCTC TGCGACAGAT 4150

GGTTCCAAGG GACCTGCCTG CACTCCTGGG GACCATAGAC CTCCAGCCTG 4200

GAGTCCCACC TTGTGCTGTT CCTGTTTCTG AGGCAGGCTT CCCACTTCCA 4250

GCCCCCCAAG CCCAGGTCCC TTGGCTCCCC CCACCCTCCT GCTCTCTCTC 4300

ACATACACAC ACACACACAC ACACAGTTTC ACACCTCCAT ATGCACACAC 4350

CTCTTCACAC AGACGTCAAT ACATTTGCCC CTCCGTCTCC TGTGCCTTGG 4400

CCCCCCAACA CTGGGCTCCC TTTCTTGTCC TCCCCCACAG TGACAACGCG 4450

CCTGCAAAAG GGAACAAGAA CCCTTCCGCC TCCAGATGGC TCCCCTGCCG 4500

CCACCCCCGA GATCAGAGTC AACCACGACC AGAGCCGGCC GGCGGGGCCA 4550

CGCCCGGGGC CACCCTCCCC AAGTCCCCAT CTCAGGTAGG GAGNGAAGTT 4600

TTGGTAGAAG GTCCCAAGCC NTCCATCNAT NTCGTCNGGG ATNGGCTGTT 4650

GTCCTCCATC CTCCCACTCC CTGTCCCCTT TCTGGCCTGG GCAGCTATGG 4700

ACCCGATGCT TTGCCCAGTG GGGGTTGGGC CTGGACCTGG GTGTCCTTTC 4750

CCCCATCCAG CTGGCATGCT TCCGGGAGGC ATCTGTGACT TGCTCGTTCC 4800

TCCCCAGCCC CCACCCCCAC TGCAGCGCCC TCCCCTTCCC TCTCCNTGGT 4850

GTTTTGTGAT GNTNGANTCT TTNTCCATNT TNTTTTCCTC CCTAGCAGAG 4900

GGTATGGGCC CTCAGCCCAC AGAGCATCCA GATNTCCAGA GTGGGCTGCC 4950

TGTCCCTNTT NNTGGCCTCT CTCTTTTAAA GGGGCCTGAG GGAGGAGCCC 5000

AAGCCAGGTG GCCATGCAGG ACCTTTAAAG GGACAGAGAG AGGAAGGGGT 5050

CAGAGGAGGG TGTGGGGTGG CTAAGGGGGC AGGTCNTGGG NTTGTGGANT 5100

GTCCTTGTGT CCACCCCGCC TGCCCCCAGC GGGCCCTCCA GGTGTANGCA 5150

GGNCCCTAGG TGCTGGCTGG CCAGGNGGGG GAGTTTTCAT AGCCGGGATC 5200

CTGCAGCTCC CGTTTTCTGC TGCCGCCCTG CTCTGCTGCT GACTAGGATA 5250

GCAGGGCTAA GGACATGGTG GGAGCCTGTC CCAAACAGCA CTTCCCCCGG 5300

CCTGGACATG GTGCCAGTGC CTTCTGTGTA TTCGTTCACT GAGTCCCCAC 5350

AACACCCCTG TGAAGCAGGC GCTGTCATCA CCTGATGCAT GAGGAAGCCC 5400

ANCGTCATGG GTGTGTGACC TGCCTGAGGT CCCCCACCTG GTGGGCAGGG 5450

GTGTGGCCTC TGCCCCATCC TGGTGCCACG CTGGCTTCCT CTGGGATACA 5500

CTCGTCTGAG CTGGGCTCCC TGTGGGCAGC CCTGTGCCCT GGGAGGTGGA 5550

AAGAGGGGCC TGCGGGAANG GAGAGGTGGG CAGGGGGAGG CTGGGGCCCG 5600

GCTGTCTCTC AACGACTGTT TGCTTCCCCA GTCTTCTCAC CAGGCCAGTG 5650

GGAGCCAGCC CCTCCCACAG TTGGCCAGTG GGCAGCCTGG GGCCTCTCTC 5700

TTCTTCGCTC TCCTTCCTCC TCTCCCCTCA CTTCTCTATC TCTTCTCTCT 5750

CCACACAGCG TTTCTGGACC GCCTGCCTCA GTGTCCCTCT CGGGGGTGGC 5800

CTGGGGTCTN GGTGTCTATG TTGGGGGGCT GGGAAGGCAN TNACTCTTCA 5850

TTTGCTGCGT CCTGCTCANT GGCCTGGGTG GGATGTGGCT GAGGTGTGAC 5900

TAACCGTGGC TTTGTCTCTG TCTGTCTCCC CCAAACCCCG TGCTCTGCTG 5950

TGCCTTCCCG CGCGGCCCCT CACCCGCCGC CGACCCACAG CTCCGGAAAG 6000

GCCCACCAGT CCCTCCGCCT CCCAAACACA CCCCGTCCAA GGAAGTCAAG 6050

CAGGAGCAGA TCCTCAGCCT GTTTGAGGAC ACGTTTGTCC CTGAGATCAG 6100

CGTGACCACC CCCTCCCAGG TCAGCCGCGG CCGCCGCGGC CCAGCTCTCC 6150

TCTCTTCCTG CCCTCTCAGG GCGTGCATGG CCTTCATCCT CTATGCTTCT 6200

GTCTCAAGAG CCAGGAATCT GGCCAGAGAG AGTGTCAGTT TCCCTCTCTC 6250

ACCCTTTGTT CCCTCCATCC ATCATCCTCC ATCATCCTCC ATCACCCATC 6300

TCTGAGCATG TACTAAGGCC AGATGCAGGG CCGCAGAGGG GAAGGTGCCG 6350

CCTCTCCCGG CGCAGCAGTT ACATCAGCAG CGCCCTCGCG ATGCAGTGGG 6400

TGCTATGGCA GAGGGGATCG GGGAGTGTGG AGGACTGTGG CTGTCAGGGA 6450

AGGCTTCCAG GGCCAGGGAG AGTTGGAAGG TCCTGGAATG GCTGAAGCAC 6500
CTGGACTTCA GCTCCCACAG CTGCTGTCAG CCCCTCGAGG GCGGGGGCAG 6550
CGGCCAGGCT GCAGGGCANA ACTGCCGGTG TGCAACACTC CCTAAGAGGC 6600
GTGGAATGCC CAGATACAGC AGGGAGCCAC CCAGGGGGGC TTGGGTCTCT 6650
CCCGACGGGC CCTTGGCTCA GCAAGGAGCC ACGCAGAGGG TCTTGGGTCT 6700
CTCCCAGTGG GCTCTTGGCT CAGCCGTGGA GGTGCCTCTG GGGAGCCCGG 6750
CCCACAGCCC CAGGTCTTAC GTCCTTCATG GTGGAGGTCG GGCTGGAGTA 6800
CCTGTGCTGG AAGCGCATCT TGCCAGTGCT GGAGTGGGCT GACGTGTTGT 6850
CAGATTTGCC CAGAGGTGGC CGGCCTTCCC CGCACTCCCC GAGAGCTGAC 6900
TGCCTCCTCA AGGTCCAGCC CTCAAGGCCT CACCTTCCTC CTGTGGGTTA 6950
GCCAAGAACC TTCCCACACA AACCTCCCCT GTTAGGAAAG CTGTCCATCC 7000
AAGCTTGTGG TGGCCTCCCA ACAAACACCT TCCACACACT CAAAAACCCT 7050
ANTGGGGANT AGTTTGGAAG GTTTTAATTT TNGGGAATTT GCCCNCTGGG 7100
AACTTGCAAA CANTGGTCCC CTGCTAAGAA AGGTTTGGGA NTGGTGGGCC 7150
TCCAACCCCC TNTGCNAAAA NNTAGGAAAT TAAAACTNAG GAACCNAAGG 7200
CNNCCGCCNC TTGCNTTTGT CCATGAAANN NNNNGCCCAC GGGCTTACCC 7250
GGNTGTGGGG TGGTGNGTAG CGTGTGTCCN TGACATGGAG GGACNGTCCC 7300
GGGCCTGCAT GGCGGGGTGC CACCTGCCGG GGCAGCACAG CGAAGGGATG 7350
GTCAGCTTTT TGGCGGATGA CCCTCCCCTC AGCACATGAC GGATATTGCT 7400
GCGTGGGTTG GCTGACTTTT ATGAGACAGG AGGGAGGGGT GTTGCTGGGG 7450
CAGGGTGGGG GCCACTGGGG AGAGATGCTG GCCGCCCGCT GGTGGGAGGC 7500
ACCTCGAGGC TGTGCACCGG CGTCCTCAGG GCTCCTTCAG AGACGGCCGG 7550
TTATGGGGCA GAGCAGTGAC CTCCCGACCC TGGGTTCCCC AGACAGGGCT 7600
GGACCTAAAG GAAAGTCAGC TGCTGGGATT GGCCCAGGGC AGGGCTTGGG 7650
GCCTAGGGCC CCTGGTTCTA GGAAGTGAGT CCACTTGGCC TGAGCTGTCT 7700
GACACCTTGG CTTGGCCATG TGGANTGCTC CACGCTTGTC CCCTGAGTGC 7750

AGGACAGCTG GTCTTCTTAG GACTGAGGAC CTTGGTNTCT CCCAATGGGC 7800

CTTCGGTTCA GNTATGGAGT GCTTNTGGGG AGCCCGGCCC ACAGCCCCAG 7850

GTCTCACATC CTTCATGGTG GAGGCCGGGT GGGAGGGCGC CCCTGTCAGT 7900

GTCCGGTGCC TGTCAAGAGT GTGTAGAGCC GGGAAGCCGC TGGCCTGGGC 7950

TGCGGGGCTG GAGTTCTTCC AGCACTGCCT GAGGGCCCCG GAGGGGAGCA 8000

CCCCGGCCAC GTCCCTCTCC TTTTAAACCT GGGCAAAGTT CTCTCTGGCC 8050

CCCAAAGGGA AGCCCCAGGT ACAAGATGGA GACCGCAGCC GAGCCAGTCC 8100

CTGCTCCTCA GAAGGCAGCT TGTGCCCTGG GCATGGATGC TGCCCCAGGG 8150

CTCACCGGAG TCATTGTCCC CGCCTGTGCC GGGGGCTCTA AGGAAGCCCC 8200

TTCCTCCCAT GCTAGTCTGG CCCAGCTTAT GGGGAGGCTT GTCCCTGTGT 8250

GGCCAGGGCC ACCGTGTCCC ATCCCTGGGG CCATGCCTGT CACATGCCTA 8300

TTCCTGGGCT CACTGGAAGG AGATCTTGGC GAGGGGCTGC TGGGAGGGGT 8350

CAGGGGCCTG CAGTTTTAAC CCAAGTGCCC CGGGTGGTTC TGAAGCCCCC 8400

GAATGTTGAA GACCCCACTT TGAAGCTTGG CTGTTGGGCT TTGTGGCTGG 8450

CTCCACTCTT TCTCCGTCCC TGGAGCTGAC GGCTGGTGGT GTCGCCAGAG 8500

AGTGACCTGC CTGTCTGGGG TGGAGGAAAA GCCAGTGTGA AGTCTCTGCC 8550

TTTGGAACTT TCCCAGTCGG GAGCACTGAG GGTGGCTGTG GCATGGTGTT 8600

ACTCTCGCCA CTGGGGGGTA GCAAGATCAG CAGAACTCTT GGCGCAGGGA 8650

GCGGAGAGGA GGTTCGGGCA TTGGTAGGGA GGGGCCCACC AGTCTGTGGA 8700

TGGTGGCGGA AGAGAGCTGG GGCCTGGTGC TGGCCCTGCG GGGTGGCGGC 8750

CACGGGCGGA CCTATGACTG GGAGTTTGAG GCGGGCACTG GGGTCGTCCT 8800

CCTGGTGTGG GCGGGAGCCT GTGCCGGGGC GCGTGGCTTT GGGCAGTGCT 8850

CCCGTGTGTG AGGTGGATGA GTTGGTGCCT GGGCTGTGTG CCAGCGTGTG 8900

TGCGTGTATG TGCGCTTGCT CTGTGCATGC GTGGTGTGTG TATGTGTGTG 8950

TGTCCACGCG TGTGCCTGTG CCTGCAGTGT CTGCCTGGGG TGAGGGCTCC 9000

CAGCTTAACA CTAACTGCTT CCTCCTCTGC TGCTGCTGCT GCTGCCAAGT 9050

TTGAGGCCCC GGGGCTTATC TCGGAGCAGG CCAGTCTGCT GGACCTGGAC 9100
TTTGACCCCC TCCCGCCCGT GACGAGCCCT GTGAAGGCAC CCACGCCCTC 9150
TGGTCAGGTT GGTTGTGCCC ACCACTGCCC ATGGGCCCAC CAGCTTCCAG 9200
GTGCCCAACC CTGGGCTCAT GTTGCCTATT GGCCACGTGA CCCCAGCTAG 9250
GCCTGGGTCA CTGCCCTTCC CCTGGCACCT CAGCCTTCAG CCCTCATCAC 9300
CTCCTGGTTG TAGGGCAGGA AGCAGCCCCT GATCAGCTGG GAGAACTCTC 9350
AGTAGGGGGT TACTGAACAC TTCCTGGCAA CTTTGTGCTC ATCGCTTGGG 9400
GCAGAAGCAT CCTGGCTTGG GGTCTTGAAG CTCCCTGAGA GGTGTCGGGA 9450
GCTCGGCCAC CTGCAAATCT TGGAGTCTAC CTGGCTCCGA GCCACTCCTG 9500
TGCCTGCTGG GCTGGATGGC CTGGGGCGAG CGGGGGTAGG GTCCCCTGGG 9550
GACTGCTTGC CGCCCTGTCT CTAACCTCTG TGCTAACTGT CCTTCTCGCC 9600
CTCACTGCTG CGCTCAGTCA ATTCCATGGG ACCTCTGGGA GGTTAAGCTG 9650
CACTCTGCTC TTTGTCCACC CCCTGGGGGA ACCACTCTTT CCCGTATGTG 9700
TCCAGGCCCA CATGATCATA GCCTGTTCAC AGGTGCATGC ACCCCACACA 9750
CCCCCCACAA GCAGGACACA CAGGCACGTG CTCACGCACA GGGAGNTGGT 9800
GAAGCCACCC GCCTCCAGCC ATTNTGNTGC TTCTCCCTCT GGCAGGCCCT 9850
TGGAAAAGGG GATCTTCGGT TTAGCTTGAG ACAGGGGTCC CCTGAGATCT 9900
GGTCCTGTTT TCACAGCCTG TGAGTGTTTG CCTCCAGACA GAAATGGGCC 9950
GGTCACCCAG GATGGACGAG TGTCCTCAGG GTGTGGGGCA GGAGGGCCTC 10000
AGGGTAGAAG GTTCTTGCCT TCTCTGAGCT TTTTGGCAGT GGGGAGCTGT 10050
TTGCGAGGAA GGGGAGAGGG GAGGAATGGA TGGTTTGAGA GAATCAGGGA 10100
AGACAGGGTG TGGCTGAGTG CCTTCTGAGA GCAGGGCCTG CAGGCAGGTG 10150
CGAGGCCATC TCACACAGCA CCATGTCACT GTCACCTGAT AGCTCAGGAC 10200
ACAGAGGCTC AGGGAAGGCT CAGTACTTGC CCAAGAACTG GTCATGGTAG 10250
AGCCAGAATT CCAACAGGGT CTCCTGGGCT CTGTCCCTGA GACCCCCTGA 10300
TACAGGCAGA GATGCTGGGA GGGGCAGGCG GGTGTGCAGG CGCCCTTGGG 10350

GCATGCGCTG GCAGCCCAGG CTCCTGGGAG CTCTGGAGGC TCCACCGCAG 10400

GATTTCCCTC TGGAGGAAGC CAGAAAGAGC CAGCCTGGTG CGAGCTGGTA 10450

GGGCCATTTT GACAAGTGGA TTTCGGTAGG TGCTGAGCTT GGGCAGCACA 10500

GTCACACCTG CCTGTCCCTT TGACAGTGGT AGGAGAGAGG ATGTGGGAGG 10550

CGGGTGGCTG GCCGGGCTCC GCTGGTACCC ACCCTGCCCC CACCAACCCC 10600

AGCCGCTGGT GACATTTTCT CTTGTCTTGT GATCCTGCCC ATTGCCTTTC 10650

CACCCCGGCC TCCCCGCCCC CTCCCTGTTC TCTCCTCGTG GCCTGTTACC 10700

AGCCCACAGA GAGTCCAGCC GGCAGCCTGC CTTCCGGGGA GCCCAGCGCT 10750

GCCGAGGGCA CCTTTGCTGT GTCCTGGCCC AGCCAGACGG CCGAGCCGGG 10800

GCCTGCCCAA GTAAGTGCCC ACCTCCAGCC CCTGTCTGGC TTGTCCCCAG 10850

TCTCTAGGGG TGCAGCATGG AAGGAGAGCC CCGAGGAGGG GTTGCAGGAG 10900

GGACCAGGCC ACCATGGATG TGAGGGTGAG GACAGGGTCC TGAGCTAGGC 10950

TGCCCCAGCA CGGGCTTGTC ACCAAGGCTG CCAAGGATGA ATGAGCGCAC 11000

TGGGCGCATC AGCCCCTCCT GCTTGCCCAC CCCAGCCCAA CCTCCCACGC 11050

AGGAAGACAT TTAGGAACAC CTACTGGTTT ATGCCAGCAC TTTCCANGTG 11100

TTGTGTCCCC CTGCCATGGA TTATNTATAG GTGCAGCAAG ATCTTGCCAC 11150

CTGCCGGTCA AGCAGGGTGG GCGGGCGGTG GCTGTGGTGG GCGAGGTCTT 11200

GGTGCCGAGA GAGCAGGGCC TGTGAGGCGG GGTTGGGGGT GGCACTATGG 11250

GGCTTGCACT GGGTTCTTCA CAGCATTGTC ACTCACATCC TTGGGCNTGC 11300

CAGCGCNTAC TATTCAGCTG CTTCCCCGGC CCAGGGCCCA GCTTGTCCAG 11350

CAGAGGCTCC CNTGGATTNT TCGAGGCACT GGGCAGCTCT AGACCNTGCT 11400

GCCAGCCAGG CGATGCCCCC GGCCCTGTTG CTTGGGTGCT GCCCTCCTGT 11450

GGCCTGTTTC CTGTGTCCTG GCTGTGTCCT GTCCTGTGTC TGACCCCAAG 11500

CCGGCATTTA TGTTGCAGCC AGCAGAGGCC TCGGAGGTGG CGGGTGGGAC 11550

CCAACCTGCG GCTGGAGCCC AGGAGCCAGG GGAGACGGCG GCAAGTGAAG 11600

CAGCCTCCGT AAGACAGCAG GGACAAAGCC CTGCCTTTTC CTCCCTGCCG 11650

CCCGCCTGCC TGTCCGGGGC TCCCCTGTGG CCCCTGATGG TGCTGGTCCA 11700
GGCCTGGCTC CTGTTGAGGA AGCTGGAGGC GGGCCGGTCT GGCACCAGGC 11750
GCAGACACCT TTCTCCCCTC CCCGCCCCTC TTCTCCTCGG TGGCCCTGGC 11800
TGTCCTTGGA CCACCTTCCC TGCTCAGCTG ACCCGTACCT CTGCCACCAG 11850
AGCTCTCTTC CTGCTGTCGT GGTGGAGACC TTCCCAGCAA CTGTGAATGG 11900
CACCGTGGAG GGCGGCTGTG GGGCCGGGCG CTTGGACCTG CCCCCAGGTT 11950
TCATGTTCAA GGTGAGCCCA CAGCCTCTGA CTGCTGCAGT CCCTCGGTGC 12000
CCTGGTGGGC AGATGACAAC CCTGAGCCTC AGGAGACTCT GTGGTTTGCC 12050
CAAAGTTGTG CAGGCGCTAC TAGGTCACTC CCAGCCAGCA AGGTGGCATC 12100
TGANCCCCAT ACAGTCCTGC TGCTTTTGAG CACTCCTGGT CTCCATACTG 12150
CCACCTGCAC CTCCCACACG CAAGGCCCGT GCTCTGTGCA GGGCTGGAGG 12200
TGGGACGGAA GGTCTGACTT GCGATCCGCA TCCTCTGCAG GTACAGGCCC 12250
ACCACGACTA CACGGCCACT GACACAGACG AGCTGCAGCT CAAGGCTGGT 12300
GATGTGGTGC TGGTGATCCC CTTCCACAAC CCTGAAGAGC AGGTGAGGGC 12350
TGGGTGGGGC CCCCACACCN CANGGGGACC ACCNNGCATC CTGGCTGCGG 12400
CTGGCACCNC CGTNGCGGAT ACNCGCCATT CAGGGGGCAG CAGAGGCCCG 12450
CGAGCACCAG GGCTCCCGCG CCAACTGCTC CTCCCCGCCC TCCACGTCGG 12500
GCTTTTTCCT CTCTCCCTCT CCTCTCCCTT CCCTTGGCCC CTCTCCTGTT 12550
AGGCCTCTCT CTCTCCCTGT CCCCATACCC GCTTCTTCCT GTAGCCTCTG 12600
CTTTCTTCTC CCCACGTCCC CCCTTTGCTC AGGCGCTCTC AGCTCTGCCT 12650
CTGTCTCTCC CCTTCTCCTC TCCTGGCAGC TGTGCCTGAG GCCTGCCTCC 12700
CTCCTGGGAC AGGATGCTTG ACCCCTCCTG CCCCGCCCAC AAGGTGCCCA 12750
CCCTGCAGCC AGCCGGAGCA CTGGTTGGGC TCATGAAGCC CCGTGTGCCG 12800
TCCCTCGAGG CGGGCCCTGC CCTGTGCACN CAGGGCCATG GGCTTCCCAG 12850
CTGTGTCCCC GGCTGAGGCT CACCCACGAT GCCTTCCAGA CCCTTCTCCT 12900
CCTGCTGTGG CTTCATGTTA ATCTCCTGGA AGTGAGGGCT CCTGTTGAGC 12950

CTGGGTGGGT GCTAAGTGTG TCCCTCCTAA GTCTTGGGAC CTCCTGGATC 13000

TGGGTCAGTT TGCCCCTCCC CAGGGGGCCT TGGAATNATN GGCAAGGAGC 13050

TTCCCCGNTG TGTAGAACCN AGCTTTGNTT GTGGGGGGTC GGTGGTGCCA 13100

TGTGGGCATC TGGTTCTTCC ACGGTTCAGC CCCTGAGCAC NTCGGGCTGT 13150

GCACAGAGGG CCTGGCCGGT TATTCCTGCT TCCAGAGAAC ATGTTTAGCC 13200

ATCAACGCTT CTGTGTGAAT AGGTTATCAG AGCGGCTGAG GGTGACAGTG 13250

GGTCTGCCTG GGTCTTGGAT GAGGCCGACC NTACTGGGGG TCCTGGGCTG 13300

GGATGTAGGG GTACCAAGTA CTTACTGAGG TCCGGGGCAG GAGGCCTGAG 13350

TGATGAGGAC CTTGTGGGCC TGGCACTGAT TTGGCCCTTT CTCNTAAGCC 13400

CCCAGGTCTT CATGGACCTC CTAGTGGGCC AGCCCTGGCT GGGTAGGATT 13450

TCAAGCAGAC TGCTACCCAG AGCCCACAGT GAGAATTGGC CTGGGGNTGC 13500

TGGAGGGGGC TCAGGGCATG AGTAGGGTCT GTGACCAGGC TGACAATGAC 13550

ACAGAGGGAA ATAACAAAGA CCCAGGTAGG CCCCAGGCAC AGCCCAGCTG 13600

CAGGGGCAGC CTCGGCCCAG CCACTGGCAG GAGTGGATGG CCATACGGCT 13650

CCCCGTGACC CACCTGGGGC CAGGGGCCTG TCAGCACTCC CAGAGAAGGC 13700

CCTGCGGGTG TCAGGATTGA AGCAAAGGGC AAGTGGAAGT TGGAGGGACT 13750

GGTGGGATGG CCCCAATCCC TCTAGAATTG TAACTTGTTG TCACTCCCAA 13800

AACTTCGTGG GGTTGTTTGA NAAGCCTGNA ATCCTGGAAG GGCTGATGTG 13850

CACATCATGC ATGCAGTGGG ACTCATCAAA ACCAGCCACG AATGGTTAGA 13900

TCCACCTGCG GACTCACAGG CTGGCTCCTG TGGTGCCTCT GGGCAGGAGC 13950

CTCAGCCAGC ANCATCAGGG AGTGCTGCCT GGAGGAGGTG TTCTCAAGGT 14000

GGGCTTGGCA GGCTGAGGCA CCAACAGCAG GAGGAGGGGC CGTCTTCCCA 14050

GCAGGTTGGA GTGGGATGCG TGCCCTGTGG GGTGGANCCC CTTGCTCATC 14100

CCTGTGCGAC CTGNTGCTCT GCCCCTCAGG ATGAAGGCTG GCTCATGGGC 14150

GTGAAGGAGA GCGACTGGAA CCAGCACAAG AAGCTGGAGA AGTGCCGTGG 14200

CGTCTTCCCC GAGAACTTCA CTGAGAGGGT CCCATGACGG CGGGGCCCAG 14250

GCAGCCTCCG GGCGTGTGAA GAACACCTCC TCCCGAAAAA TGTGTGGTTC 14300

TTTTTTTTGT TTTGTTTTCG TTTTTCATCT TTTGAAGAGC AAAGGGAAAT 14350

CAAGAGGAGA CCCCCAGGCA GAGGGGCGTT CTCCCAAAGT TTAGGTCGTT 14400

TTCCAAAGAG CCGCGTCCCG GCAAGTCCGG CGGAATTCAC CAGTGTTCCT 14450

GAAGCTGCTG TGTCCTCTAG TTGAGTTTCT GGCGCCCCTG CCTGTGCCCG 14500

CATGTGTGCC TGGCCGCAGG GCGGGGCTGG GGGCTGCCGA GCCACCATAC 14550

TTAACTGAAG CTTCGGCCGC ACCACCCGGG GAAGGGTCCT CTTTTCCTGG 14600

CAGCTGCTGT GGGTGGGGCC CAGACACCAG CCTAGCCTGC TCTGCCCCGC 14650

AGACGGTCTG TGTGCTGTTT GAAAATAAAT CTTAGTGTTC AAAACAAAAT 14700

GAAACAAAAA AAAAATGATA AAAACTCTCA GAAAACGTGT GTGTATTTGT 14750

TCTCCCTCTT CTTGTCCGTG AGTGCGGATG GAACCGTGTN ATCTGTGGCT 14800

TTCTTACTGA GATGGTCTGC CCCCGAAGGC CCGCTGCCCT GNCGCTGGTG 14850

CACCACAGGG CTTCACCCCC TGTCCCCTGG GGTTCTTAGG GGTGGTCACC 14900

TGGANGTCAN GGACTGGGGG CTTGGGTTAA GGGGCTTGGC CACCCATCTC 14950

TTGTCCCANA AATCTTGCTN ACTGCCCCCC TAACT 14985

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Phe Thr Arg His Pro Pro Val Leu Thr Pro Pro Asp Gln Glu Val
1 5 10 15

Ile

WHAT IS CLAIMED IS:

1. A mammalian nucleic acid sequence encoding a Box-dependent myc-interacting polypeptide 1 (Bin1) or a fragment thereof, isolated from cellular materials with which it is naturally associated, selected from the group consisting of:

(a) SEQ ID NO:1;

(b) SEQ ID NO:3;

(c) SEQ ID NO:6;

(d) a sequence which hybridizes to (a) - (c) under stringent conditions;

(e) an allelic variation of (a) - (c); and

(f) a fragment of (a) - (c).

2. The sequence according to claim 1 which encodes murine Bin1 SEQ ID NO:2 or a fragment thereof.

3. The sequence according to claim 1 which encodes human Bin1 SEQ ID NO:4 or a fragment thereof.

4. The sequence according to claim 1 wherein the fragment is selected from the group consisting of the SH3 domain, about nucleotides 891 to about 1412 of SEQ ID NO: 3; the NLS, about nucleotides 813-854 of SEQ ID NO: 3; and the MBD, about nucleotides 867-908 of SEQ ID NO: 3.

5. A partial murine cDNA sequence SEQ ID NO:1 which encodes a Box-dependent myc-interacting polypeptide.

6. A human cDNA sequence SEQ ID NO:3 which encodes a Box-dependent myc-interacting polypeptide.

7. A human genomic DNA sequence SEQ ID NO:6 which encodes a Box-dependent myc-interacting polypeptide.

8. A mammalian Box-dependent myc-interacting polypeptide Bin1, said polypeptide having the amino acid sequence selected from the group consisting of:

(a) murine Bin1, SEQ ID NO:2;

(b) human Bin1, SEQ ID NO:4;

(c) a fragment of (a) or (b) having Bin1 biological activity; and

(d) analogues of (a) or (b) characterized by having at least 90% homology with SEQ ID NO: 2 or SEQ ID NO:4.

9. The polypeptide according to claim 8, wherein the fragment of SEQ ID NO: 4 is selected from the group consisting of:

(a) amino acids 278-451 of SEQ ID NO: 4;

(b) amino acids 270-383 of SEQ ID NO: 4;

(c) amino acids 252-265 of SEQ ID NO: 4;

(d) amino acids 190-250 of SEQ ID NO: 4;

(e) amino acids 263-397 of SEQ ID NO: 4;

(f) amino acids 223-251 of SEQ ID NO:4;

(g) amino acids 1-222 of SEQ ID NO:4; and

(h) smaller fragments of (a) - (g) comprising about 8 amino acids.

10. A vector comprising a mammalian nucleic acid sequence encoding a Box-dependent myc-interacting polypeptide (Bin1) under the control of suitable regulatory sequences.

11. The vector according to claim 10 wherein said vector is an expression vector.

12. The vector according to claim 10 wherein said vector is a gene therapy vector.

13. A host cell transformed with the vector according to claim 10.

14. An oligonucleotide probe comprising a nucleic acid sequence selected from the group consisting of:

(a) SEQ ID NO:1;

(b) SEQ ID NO:3;

(c) SEQ ID NO:6;

(d) a nucleic acid fragment of (a) - (c) comprising at least 15 nucleotides in length,

and a detectable label which is associated with said sequence.

15. An antibody raised against a Box-dependent myc-interacting peptide (Bin1), said peptide selected from the group consisting of:

(a) SEQ ID NO:2;

(b) SEQ ID NO:4;

(c) amino acids 190-250 of SEQ ID NO: 4;

(d) amino acids 252-261 of SEQ ID NO: 4;

(e) amino acids 263-397 of SEQ ID NO: 4; and

(f) a fragment of (a) to (d) comprising about 8 amino acids.

16. The antibody according to claim 15 which recognizes an epitope in the region of amino acids 190 - 250 of SEQ ID NO:4.

17. The antibody according to claim 16 which recognizes a complex epitope in the regions of amino acids 190 - 250 and amino acids 263 - 397.

18. The antibody according to claim 15, selected from the group consisting of a chimeric antibody, a humanized antibody, a monoclonal antibody and a polyclonal antibody.

19. An anti-idiotype antibody specific for the antibody of claim 15.

20. A diagnostic reagent comprising the antibody according to claim 15 and a detectable label.

21. A therapeutic composition comprising the vector according to claim 12 and a pharmaceutically acceptable carrier.

22. A therapeutic composition comprising the anti-idiotype antibody according to claim 19 and a pharmaceutically acceptable carrier.

23. A therapeutic composition comprising a pharmaceutically acceptable carrier and a mammalian Box-dependent myc-interacting polypeptide (Bin1) selected from the group consisting of:

(a) SEQ ID NO:2;

(b) SEQ ID NO:4; and

(c) a fragment of (a) or (b) having Bin1 biological activity.

24. A method of detecting a cancer involving the c-myc oncogene or a hyperplastic disease state comprising providing a biopsy sample from a patient suspected of having said cancer or disease and incubating said sample in the presence of a diagnostic reagent according to claim 20 or an oligonucleotide probe according to claim 14.

25. A method of detecting a deficiency in Box-dependent myc-interacting peptide in a patient comprising providing a sample from a patient suspected of having said deficiency and incubating said sample in the presence of a diagnostic reagent according to claim 20.

26. A method of detecting a deficiency in Box-dependent myc-interacting peptide in a patient comprising providing a sample from a patient suspected of having said deficiency and performing the polymerase chain reaction using the oligonucleotide probe according to claim 14.

27. The use of the mammalian nucleic acid sequence according to claim 1 or the polypeptide sequence according to claim 8 in the preparation of a medicament for treating deficiencies in Box-dependent myc-interacting peptide in a patient.

28. The use of the mammalian nucleic acid sequence according to claim 1 or the polypeptide sequence according to claim 8 in the preparation of a medicament for treating a cancer or hyperplastic disease state involving the c-myc oncogene.

FIGURE 1

Partial Mouse MIP99 cDNA and Polypeptide
SEQ ID NOS. 1 and 2

```
GAG ATC AGA GTG AAC CAT GAG CCA GAG CCG GCC AGT GGG GCC TCA    45
Glu Ile Arg Val Asn His Glu Pro Glu Pro Ala Ser Gly Ala Ser
 1               5                  10                  15

CCC GGG GCT GCC ATC CCC AAG TCC CCA TCT CAG CCA GCA GAG GCC    90
Pro Gly Ala Ala Ile Pro Lys Ser Pro Ser Gln Pro Ala Glu Ala
                20                  25                  30

TCC GAG GTG GTG GGT GGA GCC CAG GAG CCA GGG GAG ACA GCA GCC   135
Ser Glu Val Val Gly Gly Ala Gln Glu Pro Gly Glu Thr Ala Ala
                35                  40                  45

AGT GAA GCA ACC TCC AGC TCT CTT CCG GCT GTG GTG GTG GAG ACC   180
Ser Glu Ala Thr Ser Ser Ser Leu Pro Ala Val Val Val Glu Thr
                50                  55                  60

TTC TCC GCA ACT GTG AAT GGG GCG GTG GAG GGC AGC GCT GGG ACT   225
Phe Ser Ala Thr Val Asn Gly Ala Val Glu Gly Ser Ala Gly Thr
                65                  70                  75

GGA CGC TTG GAC CTG CCC CCG GGA TTC ATG TTC AAG GTT CAA GCC   270
Gly Arg Leu Asp Leu Pro Pro Gly Phe Met Phe Lys Val Gln Ala
                80                  85                  90

CAG CAT GAT TAC ACG GCC ACT GAC ACT GAT GAG CTG CAA CTC AAA   315
Gln His Asp Tyr Thr Ala Thr Asp Thr Asp Glu Leu Gln Leu Lys
                95                 100                 105

GCT GGC GAT GTG GTG TTG GTG ATT CCT TTC CAG AAC CCA GAG GAG   360
Ala Gly Asp Val Val Leu Val Ile Pro Phe Gln Asn Pro Glu Glu
               110                 115                 120

CAG GAT GAA GGC TGG CTC ATG GGT GTG AAG GAG AGC GAC TGA       402
Gln Asp Glu Gly Trp Leu Met Gly Val Lys Glu Ser Asp
               125                 130
```

FIGURE 2A

Human MIP99 cDNA and Polypeptide
SEQ ID NOS. 3 and 4

```
GAATTCCGTG CTGGTTGAGC TTGCTCATCT CCTTGTGGAA GTTTTCCTCC          50

AGGCCCGCG ATG CTC TGG AAC GTG GTG ACG GCG GGA AAG ATC GCC        95
          Met Leu Trp Asn Val Val Thr Ala Gly Lys Ile Ala
           1               5                  10

AGC AAC GTG CAG AAG AAG CTC ACC CGC GCG CAG GAG AAG GTT CTC    140
Ser Asn Val Gln Lys Lys Leu Thr Arg Ala Gln Glu Lys Val Leu
        15                  20                  25

CAG AAG CTG GGG AAG GCA GAT GAG ACC AAG GAT GAG CAG TTT GAG    185
Gln Lys Leu Gly Lys Ala Asp Glu Thr Lys Asp Glu Gln Phe Glu
        30                  35                  40

CAG TGC GTC CAG AAT TTC AAC AAG CAG CTG ACG GAG GGC ACC CGG    230
Gln Cys Val Gln Asn Phe Asn Lys Gln Leu Thr Glu Gly Thr Arg
        45                  50                  55

CTG CAG AAG GAT CTC CGG ACC TAC CTG GCC TCC GTC AAA GCC ATG    275
Leu Gln Lys Asp Leu Arg Thr Tyr Leu Ala Ser Val Lys Ala Met
        60                  65                  70

CAC GAG GCT TCC AAG AAG CTG AAT GAG TGT CTG CAG GAG GTG TAT    320
His Glu Ala Ser Lys Lys Leu Asn Glu Cys Leu Gln Glu Val Tyr
        75                  80                  85

GAG CCC GAT TGG CCC GGC AGG GAT GAG GCA AAC AAG ATC GCA GAG    365
Glu Pro Asp Trp Pro Gly Arg Asp Glu Ala Asn Lys Ile Ala Glu
        90                  95                  100

AAC AAC GAC CTG CTG TGG ATG GAT TAC CAC CAG AAG CTG GTG GAC    410
Asn Asn Asp Leu Leu Trp Met Asp Tyr His Gln Lys Leu Val Asp
        105                 110                 115

CAG GCG CTG CTG ACC ATG GAC ACG TAC CTG GGC CAG TTC CCC GAC    455
Gln Ala Leu Leu Thr Met Asp Thr Tyr Leu Gly Gln Phe Pro Asp
        120                 125                 130

ATC AAG TCA CGC ATT GCC AAG CGG GGG CGC AAG CTG GTG GAC TAC    500
Ile Lys Ser Arg Ile Ala Lys Arg Gly Arg Lys Leu Val Asp Tyr
        135                 140                 145

GAC AGT GCC CGG CAC CAC TAC GAG TCC CTT CAA ACT GCC AAA AAG    545
Asp Ser Ala Arg His His Tyr Glu Ser Leu Gln Thr Ala Lys Lys
        150                 155                 160
```

FIGURE 2B

AAG GAT GAA GCC AAA ATT GCC AAG GCC GAG GAG GAG CTC ATC AAA 590
Lys Asp Glu Ala Lys Ile Ala Lys Ala Glu Glu Glu Leu Ile Lys
165 170 175

GCC CAG AAG GTG TTT GAG GAG ATG AAT GTG GAT CTG CAG GAG GAG 635
Ala Gln Lys Val Phe Glu Glu Met Asn Val Asp Leu Gln Glu Glu
180 185 190

CTG CCG TCC CTG TGG AAC AGC CGC GTA GGT TTC TAC GTC AAC ACG 680
Leu Pro Ser Leu Trp Asn Ser Arg Val Gly Phe Tyr Val Asn Thr
195 200 205

TTC CAG AGC ATC GCG GGC CTG GAG GAA AAC TTC CAC AAG GAG ATG 725
Phe Gln Ser Ile Ala Gly Leu Glu Glu Asn Phe His Lys Glu Met
210 215 220

AGC AAG CTC AAC CAG AAC CTC AAT GAT GTG CTG GTC GGC CTG GAG 770
Ser Lys Leu Asn Gln Asn Leu Asn Asp Val Leu Val Gly Leu Glu
225 230 235

AAG CAA CAC GGG AGC AAC ACC TTC ACG GTC AAG GCC CAG CCC AGA 815
Lys Gln His Gly Ser Asn Thr Phe Thr Val Lys Ala Gln Pro Arg
240 245 250

AAG AAA AGT AAA CTG TTT TCG CGG CTG CGC AGA AAG AAG AAC AGT 860
Lys Lys Ser Lys Leu Phe Ser Arg Leu Arg Arg Lys Lys Asn Ser
255 260 265

GAC AAC GCG CCT GCA AAA GGG AAC AAG AGC CCT TCG CCT CCA GAT 905
Asp Asn Ala Pro Ala Lys Gly Asn Lys Ser Pro Ser Pro Pro Asp
270 275 280

GGC TCC CCT GCC GCC ACC CCC GAG ATC AGA GTC AAC CAC GAG CCA 950
Gly Ser Pro Ala Ala Thr Pro Glu Ile Arg Val Asn His Glu Pro
285 290 295

GAG CCG GCC GGC GGG GCC ACG CCC GGG GCC ACC CTC CCC AAG TCC 995
Glu Pro Ala Gly Gly Ala Thr Pro Gly Ala Thr Leu Pro Lys Ser
300 305 310

CCA TCT CAG CCA GCA GAG GCC TCG GAG GTG GCG GGT GGG ACC CAA 1040
Pro Ser Gln Pro Ala Glu Ala Ser Glu Val Ala Gly Gly Thr Gln
315 320 325

CCT GCG GCT GGA GCC CAG GAG CCA GGG GAG ACT TCT GCA AGT GAA 1085
Pro Ala Ala Gly Ala Gln Glu Pro Gly Glu Thr Ser Ala Ser Glu
330 335 340

FIGURE 2C

```
GCA GCC TCC AGC TCT CTT CCT GCT GTC GTG GTG GAG ACC TTC CCA   1130
Ala Ala Ser Ser Ser Leu Pro Ala Val Val Val Glu Thr Phe Pro
        345                 350                 355

GCA ACT GTG AAT GGC ACC GTG GAG GGC GGC AGT GGG GCC GGG CGC   1175
Ala Thr Val Asn Gly Thr Val Glu Gly Gly Ser Gly Ala Gly Arg
        360                 365                 370

TTG GAC CTG CCC CCA GGT TTC ATG TTC AAG GTA CAG GCC CAG CAC   1220
Leu Asp Leu Pro Pro Gly Phe Met Phe Lys Val Gln Ala Gln His
        375                 380                 385

GAC TAC ACG GCC ACT GAC ACA GAC GAG CTG CAG CTC AAG GCT GGT   1265
Asp Tyr Thr Ala Thr Asp Thr Asp Glu Leu Gln Leu Lys Ala Gly
        390                 395                 400

GAT GTG GTG CTG GTG ATC CCC TTC CAG AAC CCT GAA GAG CAG GAT   1310
Asp Val Val Leu Val Ile Pro Phe Gln Asn Pro Glu Glu Gln Asp
        405                 410                 415

GAA GGC TGG CTC ATG GGC GTG AAG GAG AGC GAC TGG AAC CAG CAC   1355
Glu Gly Trp Leu Met Gly Val Lys Glu Ser Asp Trp Asn Gln His
        420                 425                 430

AAG AAG CTG GAG AAG TGC CGT GGC GTC TTC CCC GAG AAC TTC ACT   1400
Lys Lys Leu Glu Lys Cys Arg Gly Val Phe Pro Glu Asn Phe Thr
        435                 440                 445

GAG AGG GTC CCA TGACGGCGGG GCCCAGGCAG CCTCCGGGCG TGTGAAGAAC   1452
Glu Arg Val Pro
        450

ACCTCCTCCC GAAAAATGTG TGGTTCTTTT TTTTGTTTTG TTTTCGTTTT   1502

TCATCTTTTG AAGAGCAAAG GGAAATCAAG AGGAGACCCC CAGGCAGAGG   1552

GGCGTTCTCC CAAAGTTTAG GTCGTTTTCC AAAGAGCCGC GTCCCGGCAA   1602

GTCCGGCGGA ATTCACCAGT GTTCCTGAAG CTGCTGTGTC CTCTAGTTGA   1652

GTTTCTGGCG CCCCTGCCTG TGCCCGCATG TGTGCCTGGC CGCAGGGCGG   1702

GGCTGGGGGC TGCCGAGCCA CCATACTTAA CTGAAGCTTC GGCCGCACCA   1752

CCCGGGGAAG GGTCCTCTTT TCCTGGCAGC TGCTGTGGGT GGGGCCCAGA   1802

CACCAGCCTA GCCTGCTCTG CCCCGCAGAC GGTCTGTGTG CTGTTTGAAA   1852

ATAAATCTTA GTGTTCAAAA CAAAATGAAA CAAAAAAAAA AATGATAAAA   1902

ACTCTCAAAA AAACAAGGAA TTC                                1925
```

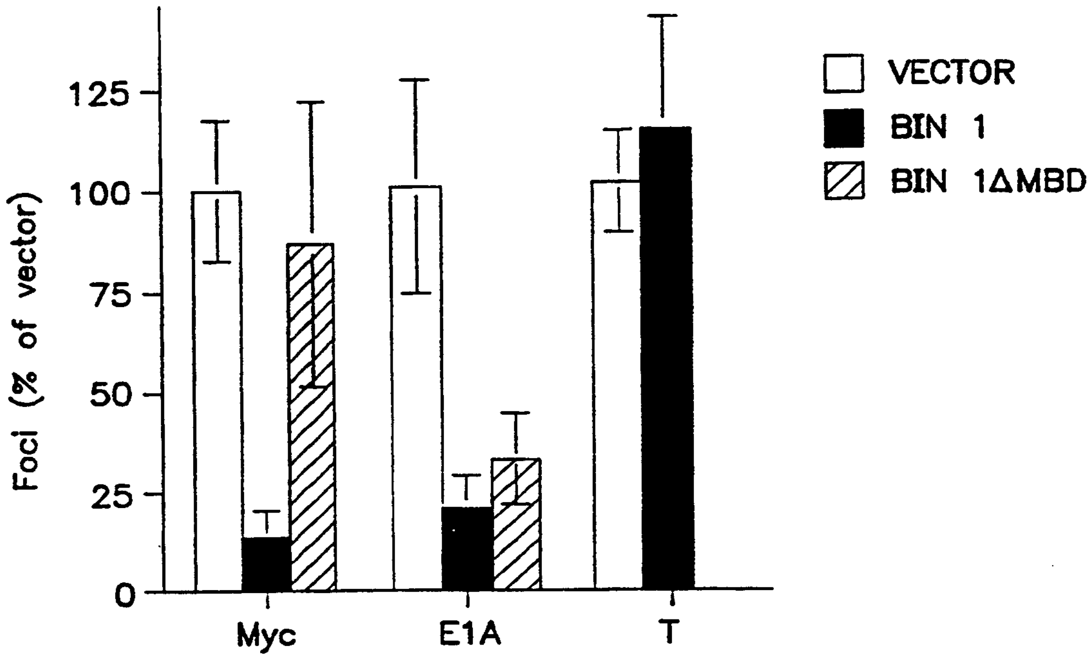
125
100
75
50
25
0
Foci (% of vector)
Myc
E1A
T
VECTOR
BIN 1
BIN 1ΔMBD

FIG. 3A

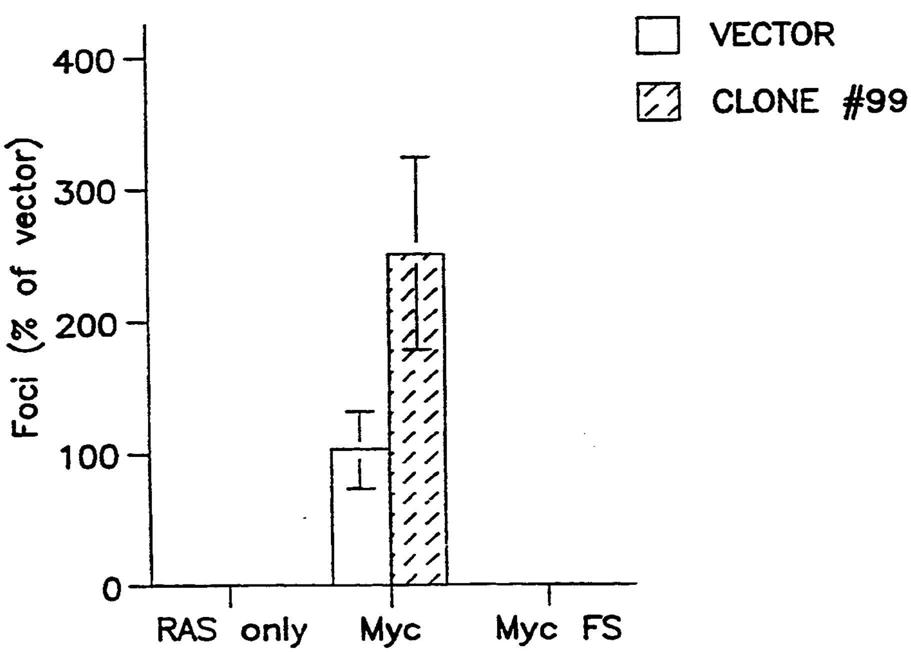
400
300
200
100
0
Foci (% of vector)
RAS only
Myc
Myc FS
VECTOR
CLONE #99

FIG. 3B

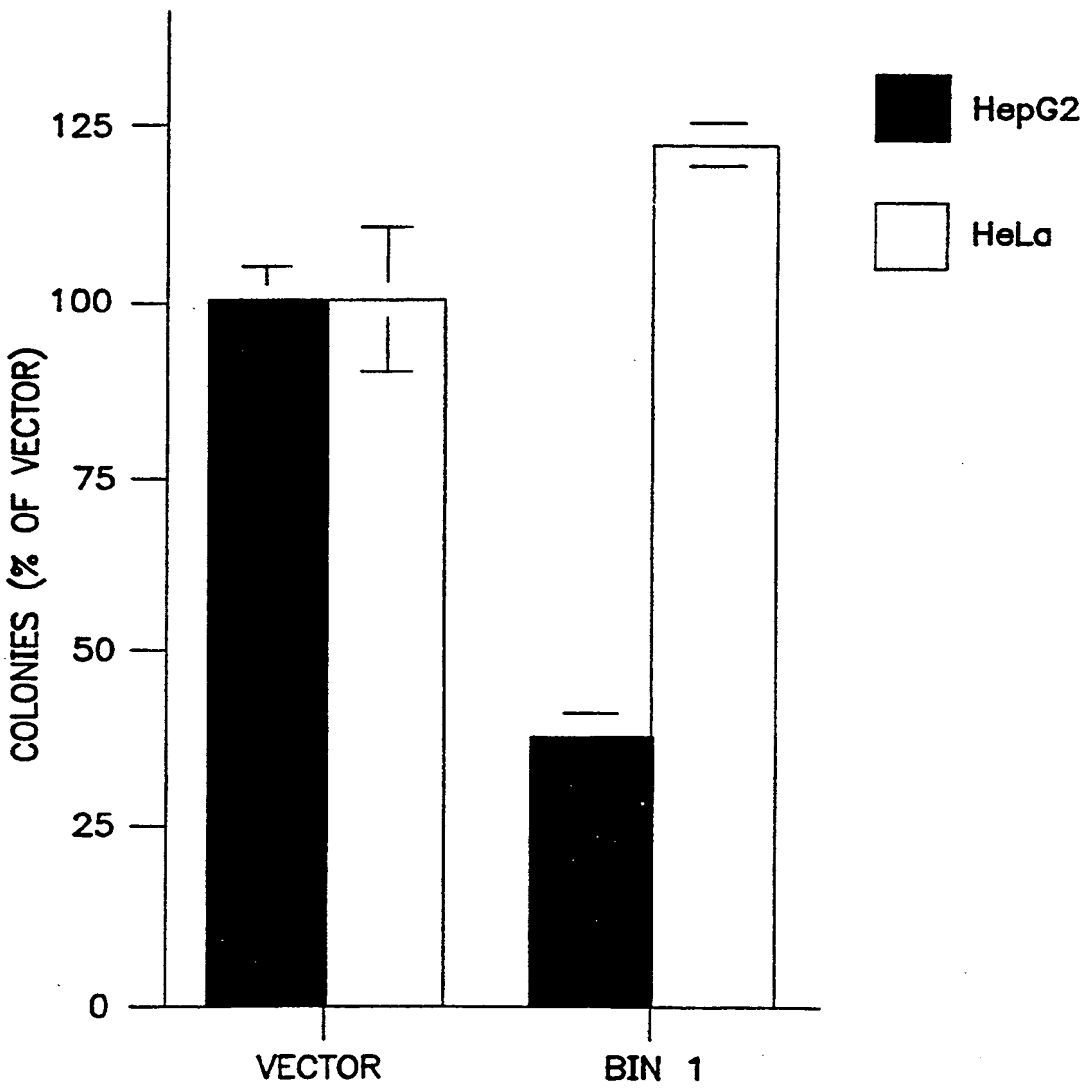


FIG. 4

# INTERNATIONAL SEARCH REPORT

| International application No. |
|---|
| PCT/US96/06231 |

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/130.1, 133.1; 435/6, 7.1, 7.8, 69.1, 172.1, 240.1, 320.1; 514/2, 44; 536/23.1, 23.5, 24.31

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS,STN,MPSRCH
sequences, author search, bin1, mip, myc, interacting, protein, antibody,box dependent

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y,P | NEGOREV, D. et al. The Bin1 Gene Localizes to Human Chromosome 2q14 by PCR Analysis of Somatic Cell Hybrids and Fluorescence in Situ Hybridization. GENOMICS. 15 April 1996, Vol. 33, No.2, pages 329-331, see entire document | 1-14, 21, 23,24, 26-28 |
| A | COLE, M. et al. The myc Oncogene: Its Role in Transformation and Differentiation. Ann. Rev. Genet. 1986, Vol. 20, pages 361-384 | all |
| A | FEILDS, S. et al. A novel genetic system to detect protein-protein interactions. NATURE. 20 July 1989, Vol. 340, pages 245-246. | all |

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 20 JUNE 1996 | 04 SEP 1996 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230 | Authorized officer<br>SEAN McGARRY<br>For<br>Telephone No. (703) 308-0196 |

Form PCT/ISA/210 (second sheet)(July 1992)★

INTERNATIONAL SEARCH REPORT

| International application No. |
|---|
| PCT/US96/06231 |

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | MARK, G. et al. "Humanization of Monoclonal Antibodies". The Handbook of Experimental Pharmacology, Vol. 113, Springer-Verlag. 1994, pages 105-134. | 15-20 |
| X,P | Databases EMBL/Genebank/DDBJ on MPSRCH. Accesion No Z24792. Auffray et al, 30 July 1993. | 1-4, 14 |
| X | Databases EMBL/GeneBank/DDBJ on MPSRCH. Accesion No. Z24784. Auffray et al, 30 July 1993. | 1-4, 14 |
| X | Databases EMBL/GeneBank/DDBJ on MPSRCH. Accesion No. F00405. Auffray et al, 07 March 1995. | 1-4, 14 |
| X | Database IMAGE Consortium, LLNL on MPSRCH. Accesion No. R34418. Hillier et al, 02 May 1995. | 1-4, 14 |
| X | Database Genexpress on MPSRCH. Accesion No. Z28487. Auffray et al, 09 December 1993. | 1-4, 14 |

## INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/06231

**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

A61K 39/395; C12Q 1/68; G01N 33/53; C12N 15/00, 15/09, 15/11, 15/64

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

424/130.1, 133.1; 435/6, 7.1, 7.8, 69.1, 172.1, 240.1, 320.1; 514/2, 44; 536/23.1, 23.5, 24.31

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1-7, 10-14, 21, 24 and 26-28, drawn to a mammalian nucleic acid sequence encoding a Box dependent myc-interacting polypeptide and methods of use thereof.
Group II, claims 8-9 and 23, drawn to a mammalian Box-dependent myc-interacting polypeptide and methods of use thereof.
Group III, claims 15-20, 22 and 25, drawn to an antibody raised against a Box-dependent myc-interacting polypeptide and uses thereof.
The inventions listed as Groups I, II and III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The special technical feature of Group I is drawn to nucleic acids that code for a Box dependent myc-interacting polypeptide and methods of use thereof and is not present in either groups II or III. The special technical feature of Group II drawn to a mammalian Box-dependent myc-interacting polypeptide and methods of use thereof and is not present in Group III, wherein the special technical feature is drawn to antibodies raised against the polypeptide of group II and uses thereof and anti-idiotype antibodies raised to said antibodies. These anti-idiotype antibodies share no special technical feature with Group I or II.